THE RELATIONSHIP BETWEEN BODY MASS INDEX AND BREAST CANCER RECURRENCE/PROGRESSION AND BREAST CANCER-SPECIFIC DEATH

Cedomir Tosevski, BA.

Submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Applied Health Sciences
(Health Science)

Supervisor: C. Martin Tammemägi, PhD

The Faculty of Applied Health Sciences
Brock University
St. Catharines, ON.

ACKNOWLEDGEMENTS

The task of the leader is to get his people from where they are to where they have not been

- Henry Kissinger

I would like to express my uttermost gratitude to the leader of this mission, Dr. C. Martin Tammemägi. No words can ever describe your contribution and involvement during this journey. You are my role model in life and in my academic career. Without your patience, leadership, guidance, support and effort I would not be where I am today. Thank you.

I am extremely grateful to have Drs. Brent Faught and Jian Liu on my committee. These dedicated scholars have been very useful in my academic development and success. Thank you for your time and feedback. I would also like to send my regards to Dr. Margie Parthimos (my external examiner) for her active involvement and feedback.

Mom and dad you have always stood by my side and pointed me in the right direction in life. You have been there during good and bad times and always assured me that I had a shoulder to cry on. You placed your faith in me and supported my decisions. I dedicate this thesis to you. Without your guidance I would not be half the man I am today. Thank you.

Grandma and grandpa thank you for your support and generosity. To all my friends, thank you for all the laughs and Lauren thank you for your encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AJCC – American Joint Committee on Cancer

BC – Bias-Corrected

BMI – Body Mass Index

$BMI_{3splines}$ – Body Mass Index Three Splines

CI – Confidence Interval

DCIS – Ductal Carcinoma in Situ

ER – Estrogen Receptor

HFHS – Henry Ford Health System

HR – Hazard Ratio

HRT – Hormone Replacement Therapy

IBC – Inflammatory Breast Cancer

IDC – Invasive/Infiltrating Ductal Carcinoma

IGF – Insulin Growth Factor

IGFBG – Insulin Growth Factor Binding Globulin

ILC – Invasive/Infiltrating Lobular Carcinoma

KBCR – Korean Breast Cancer Registry

LCIS – Lobular Carcinoma in Situ

LRT – Likelihood Ratio Test

OR – Odds Ratio

PR – Progesterone Receptor

QI – Quetelet Index

ROC AUC – Receiver Operator Characteristic Area under the Curve

RR – Relative Risk

RSR – Relative Survival Ratio

SBR – Scarff-Bloom-Richardson Histopathologic Grading System

SE – Standard Error

SEER – Surveillance, Epidemiology, and End Results

SES – Socioeconomic Status

SHBG – Sex Hormone-Binding Globulin

SNUHBCC – Seoul National University Hospital Breast Cancer Center

TNM – Tumor-Node-Metastasis Classification System

WHO – World Health Organization

ABSTRACT

THE RELATIONSHIP BETWEEN BODY MASS INDEX AND BREAST CANCER RECURRENCE/PROGRESSION AND BREAST CANCER-SPECIFIC DEATH

Cedomir Tosevski

Master of Science

Faculty of Applied Health Sciences

Brock University

2010

The aim of this study was to describe the nonlinear association between body mass index (BMI) and breast cancer outcomes and to determine whether BMI improves prediction of outcomes. A cohort of 906 breast cancer patients diagnosed at Henry Ford Health System, Detroit (1985-1990) were studied. The median follow-up was 10 years. Multivariate logistic regression was used to model breast cancer recurrence/progression and breast cancer-specific death. Restricted cubic splines were used to model nonlinear effects. Receiver operator characteristic areas under the curves (ROC AUC) were used to evaluate prediction. BMI was nonlinearly associated with recurrence/progression and death (p= 0.0230 and 0.0101). Probability of outcomes increased with increase or decrease of BMI away from 25. BMI splines were suggestive of improved prediction of death. The ROC AUCs for nested models with and without BMI were 0.8424 and 0.8331 (p= 0.08). If causally associated, modifying patients BMI towards 25 may improve outcomes.

CHAPTER 1: INTRODUCTION

1.1 The Impact of Breast Cancer among Canadian Women

On average, 445 Canadian women will be diagnosed with breast cancer and 100 Canadian women will die of breast cancer every week. Canadian women have a 1 in 9 lifetime cumulative risk of developing breast cancer and a 1 in 28 chance of dying from breast cancer-specific death (based on life tables standardized to the Canadian census population from 1991 up to 85 years of age).[1] These estimates are representative of an age-standardized incidence rate of 102 per 100,000, the highest among all cancers, as well as, an age-standardized mortality rate of 21 per 100,000, second only to lung cancer. *Figures 1* and *2* display trends in cancer incidence and mortality over the past 30 years.[1] Much of the long-term increase in incidence may be attributed to the increased prevalence of mammography screening and increased utilization of population based screening programs. Additionally, changes in reproductive patterns such as delayed childbearing and having fewer children, the rising prevalence rate of obesity and postmenopausal hormone use (i.e. hormone replacement therapy (HRT)) may have contributed to the increase in breast cancer incidence. Meanwhile, reduced breast cancer mortality rates have been attributed to improvements in treatment and mammography screening (detecting those women diagnosed with breast cancer at an earlier stage of disease).[2, 3]

In Ontario women, according to Canadian Cancer Statistics 2010, breast cancer is the most frequently diagnosed cancer, followed by colorectal cancer and lung cancer.[1] Lung cancer is the leading cause of cancer death, followed by breast cancer and

colorectal cancer.[1] In 2010, an estimated 8,900 women will be diagnosed with breast cancer and approximately 2,100 women will die of breast cancer.[1]

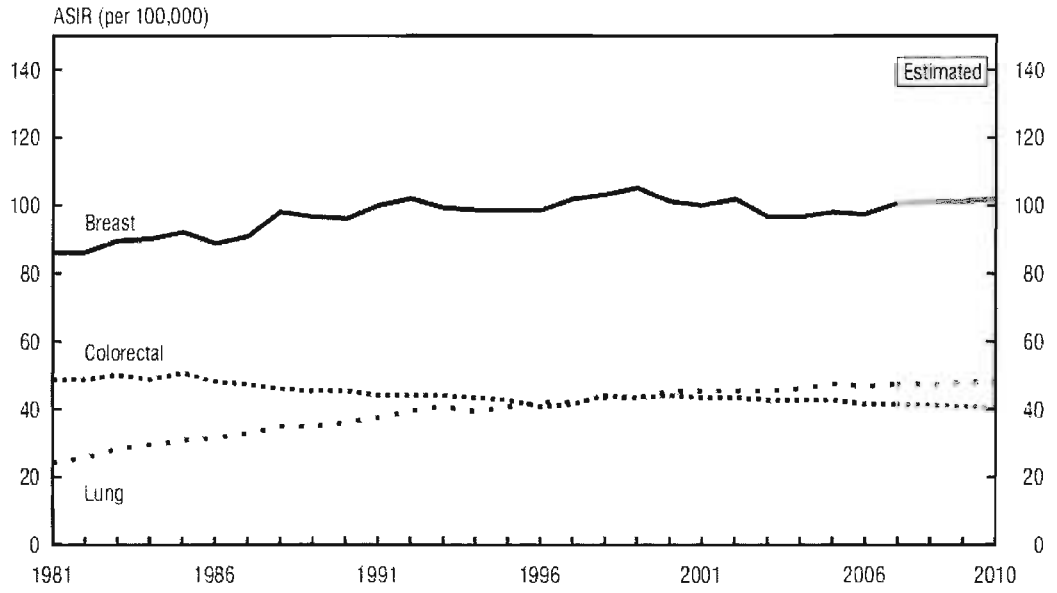


Figure 1. **Age-Standardized Incidence Rates for Selected Cancers, Females, Canada, 1981-2010.[1]**

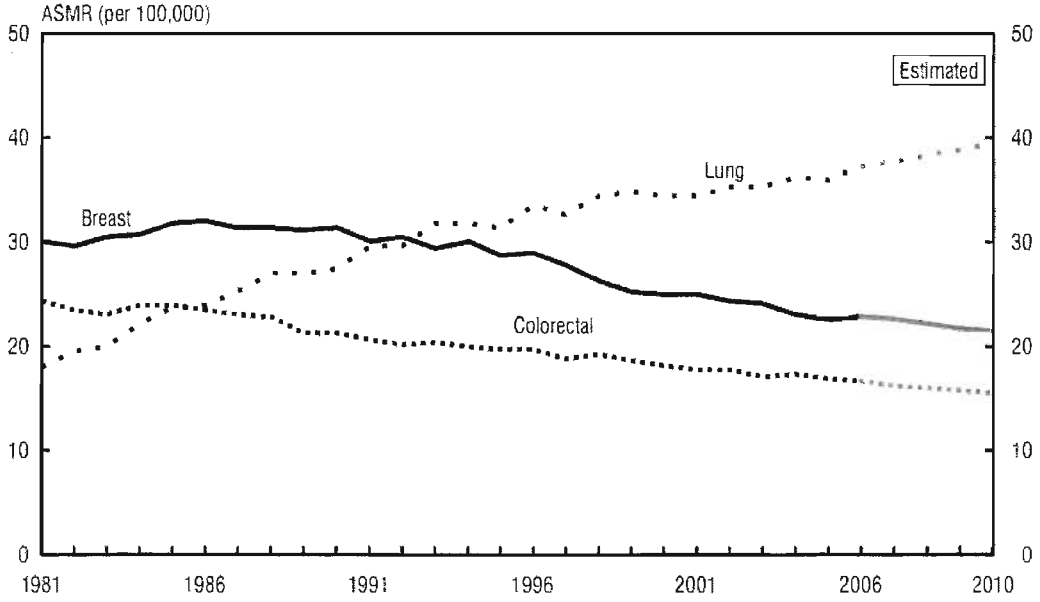


Figure 2. **Age-Standardized Mortality Rates for Selected Cancers, Females, Canada, 1981-2010.[1]**

1.2 Why Study Obesity and Breast Cancer Outcomes?

In the last decade the prevalence of obesity has nearly doubled. This increase in prevalence is evident in the US and Canada, with the highest prevalence of obesity reported in the US.[4-6] Studies show that body mass index (BMI) is a risk factor for breast cancer, as well as, a prognostic factor for breast cancer recurrence/progression and overall survival. The risk of breast cancer is modified by menopausal status and is therefore different in postmenopausal and premenopausal women.[7] Generally, postmenopausal women with a high BMI are at an increased risk of breast cancer whereas premenopausal women with a high BMI are at a lower risk of developing breast cancer. Studies evaluating the prognostic relationship between BMI and breast cancer outcomes show that women with a high BMI at breast cancer diagnosis have shorter overall survival and a higher risk of breast cancer recurrence/progression.[4, 8-10]

1.3 Gaps in the Literature

Although past studies evaluated the relationship between BMI and breast cancer outcomes, results have been inconsistent and limited. Major emphasis was placed on the role of high BMI and literature failed to simultaneously assess the impact of low BMI on breast cancer outcomes. In addition, several investigations focused on all cause mortality, which would hide relationships specific to breast cancer death. Recent literature has analyzed BMI but only in broad categorical groups or as linear effects. Greenberg and colleagues, Katoh and colleagues, Menon and colleagues and Carmichael and colleagues modeled BMI in broad categorical groups and as a result did not find a significant association between BMI and breast cancer recurrence and overall survival.[11-14] Katoh and colleagues used a BMI > 27 kg/m^2 vs. a BMI of < 27 kg/m^2 whereas Greenberg and

colleagues used a BMI of > 27 kg/m^2 vs. a BMI of < 20 kg/m^2 to define high and low BMI.[11, 12] Modeling BMI in broad categorical groups may have accounted for the null findings demonstrated by these authors. Finally, studies have yet to assess prediction using sophisticated statistical techniques such as bootstrapping for internal validation and assessment of discrimination and calibration, leaving a large gap in knowledge.

1.4 Response to Gaps in the Literature

In order to address the gaps in the literature the following research questions are addressed:

(1) Determine whether BMI near or at breast cancer diagnosis is independently associated with breast cancer recurrence/progression and breast cancer-specific death, after covariate adjustment?

(2) Evaluate whether BMI is nonlinearly associated with breast cancer recurrence/progression and breast cancer-specific death?

(3) Determine whether properly modeled BMI improves the ability to predict breast cancer recurrence/progression and breast cancer-specific death?

In this study, BMI was modeled using restricted cubic splines. Restricted cubic splines are higher ordered polynomials used in exploring complicated nonlinear relationships. They are versatile and can fit highly curving shapes. To determine whether BMI was independently associated with breast cancer outcomes, BMI was adjusted for well-established covariates, which were chosen *a priori* and from past studies, such as age, stage, estrogen receptor/progesterone receptor hormone (ER/PR) status, surgery, radiation therapy and chemotherapy. To determine whether BMI improves the ability to predict breast cancer outcomes, prediction was evaluated using bootstrapping for internal

validation and by assessing discrimination and calibration. Bootstrapping estimates the sampling distribution of an estimator by resampling with replacement from the original sample.[15] Discrimination measures the models ability to separate those women who will experience a breast cancer recurrence/progression or die from breast cancer from those who will not by calculating the proportion of pairs of women in which the woman with breast cancer recurrence/progression or breast cancer-specific death has a higher predicted risk than the woman without breast cancer recurrence/progression or breast cancer-specific death.[15] Finally, calibration assesses how well the predicted probabilities match the observed probabilities.[15, 16]

1.5 Conclusions

This study is important for a multiplicity of reasons. The purpose of the association study was to determine if BMI near or at breast cancer diagnosis was associated with breast cancer recurrence/progression and breast cancer-specific death. Considering that BMI and breast cancer outcomes are related, this study would allow further investigation into possible etiologic or causal mechanisms. In addition, the results of this study may aid in the formulation of intervention programs such as nutritional and exercise training programs for women diagnosed with breast cancer as these programs may be beneficial to the patient (and the health care system at large). On the other hand, prediction models are used in clinical settings by physicians and health policy makers in order to generate predictions regarding the probability/risk of a patient surviving breast cancer. Therefore, models predicting breast cancer outcomes need to be able to discriminate correctly, while at the same time be able to optimize calibration and goodness-of-fit.

CHAPTER 2: REVIEW OF THE LITERATURE

2.0 Overview

The current chapter provides the rationale and basis for this study. The chapter begins with an overview of the biology and etiology of breast cancer. Then, age-standardized incidence and mortality rates are briefly reported for the US and Canada. Provided is a discussion of the breast cancer risk factors, pertinent prognostic factors, as well as, the role of breast cancer prognostic models in clinical and public health settings. The chapter ends with a focus on body mass index (BMI) as a relevant prognostic factor for breast cancer recurrence/progression and breast cancer-specific death. Particular attention is given to modeling BMI using restricted cubic splines and its nonlinear association with breast cancer outcomes.

2.1 Defining Cancer

Cancer is a term used for diseases in which abnormal cells uncontrollably mutate, grow and proliferate within the body.[17] All cancers are initiated within cells. Under normal conditions, cells grow and divide in a controlled manner in order to produce additional cells, as cells are required for bodily functions. Eventually, these cells become old or damaged and as a result these cells die and are replaced with new cells. However, this process can sometimes be disrupted. The genetic material of a cell can become damaged or altered, resulting in alterations in DNA base-pair sequences which affect normal cell growth and division. As a result, cells are alive when they should be dead but because they are damaged or mutated, they do not function normally and new cells start to form when the body does not require these extra cells. Therefore, these excess cells are capable of forming a mass or tissue referred to as a tumor.[17]

2.1.1 Breast Cancer Categories

Cancer is a complex disease that encompasses multiple disease categories. The five main categories include:[17]

(1) ***Carcinoma:*** Cancer that begins in epithelial cells or cells lining the mesodermal origin. In addition, carcinomas have the ability to invade surrounding tissues and organs and metastasize to lymph nodes or other sites.

(2) ***Sarcoma:*** Cancer that is initiated in bone, cartilage, fat, muscle, blood vessels and other connective or supportive tissues of mesodermal origin.

(3) ***Leukemia:*** Cancer that initiates in blood-forming tissue such as bone marrow and causes large numbers of abnormal blood cells to be produced and to enter the blood.

(4) ***Lymphoma and myeloma:*** Cancers that start in the cells of the immune system.

(5) ***Central nervous system cancers:*** Cancers arising in the tissues of the brain and spinal cord.

Nearly all breast cancers are classified as carcinomas because the cancer begins within the epithelial cells of the breast.

Breast tissue is made up of glands for the production of milk (lobules) and the ducts that connect the lobules to the nipple (*Figure 3*). The remainder of the breast is composed of fatty, connective and lymphatic tissues.[18]

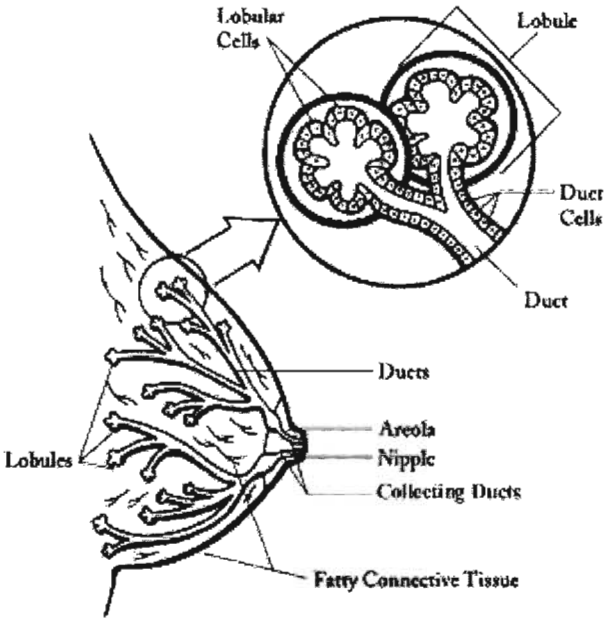


Figure 3. **The Breast.**[19]

2.2 Types of Breast Cancer

Two types of masses affect breast tissue: benign or invasive/infiltrating. Benign masses are not cancerous, they do not uncontrollably grow or spread and are therefore not life threatening. In contrast to benign masses, majority of breast cancers are potentially invasive. These tumors are deemed cancerous and have the ability to metastasize or spread from the lobules or ducts of the breast into surrounding tissue or regional lymph nodes or distant tissues or lymph nodes.[17, 18]

2.2.1 Ductal and Lobular Carcinoma in Situ (DCIS and LCIS)

DCIS is the most common type of non-invasive breast cancer. DCIS and LCIS are termed in situ because the cancer is confined to either the ducts or lobules.[18] DCIS is only evident in the ducts of the breast because it does not spread through the basement membrane of the epithelium and into surrounding breast tissue. On the other hand, LCIS

is initiated in the lobules and does not penetrate the lobular walls. DCIS is the most common type of non-invasive breast cancer.[18, 20]

2.2.2 Invasive/Infiltrating Ductal and Lobular Carcinoma (IDC and ILC)

IDC is initiated within the ducts of the breast. The cancer eventually spreads through the basement membrane of the epithelium and invades the breast tissue. Once the cancer penetrates the basement membrane of the epithelium it is able to metastasize to other parts of the body. IDC accounts for 8 out of 10 invasive breast cancers.[19] On the other hand, ILC penetrates through the lobular epithelium's basement membrane and then can metastasize to other parts of the body. Approximately, 1 out of 10 breast cancers are classified as ILC.[19]

2.2.3 Inflammatory Breast Cancer (IBC)

IBC is characterized as rapidly developing (< 3 months) signs and symptoms of diffuse erythema, peau d'orange (orange skin) and increasing size of the breast(s) with or without evidence of extensive dermal lymphatic invasion on core biopsy specimens.[8, 19] IBC is a rare type of breast cancer that can grow and metastasize suddenly, even at an early stage.[21, 22] IBC is likely to develop when breast cancer cells block the lymph vessels which are responsible for removing fluids, bacteria and waste products from the breast tissue resulting in inflammation of the breast.[21] IBC accounts for 1-3% of breast cancers diagnosed in the US.[19] Women diagnosed with IBC have a lower overall survival and an increased risk of breast cancer-specific death.[22]

2.3 Breast Cancer Staging

2.3.1 Brief History

In 1904, Steinthal, a German physician, proposed the division of breast cancer into three distinct stages: (1) Stage I- small tumors that were localized to the breast, (2) Stage II- large tumors involving axillary lymph nodes, and (3) Stage III- tumors that invaded tissue around the breast. Later, Steinthal's staging system was refined by Greenough, who based his classification on microscopic evaluation of breast cancer specimens.

Following Greenough's revision, the tumor-node-metastasis (TNM) classification system for breast cancer staging was developed by Pierre Denoix beginning in 1942.[23] The TNM principal objective, as displayed in *Table 1,* was to classify cancer based on major morphological attributes of malignant tumors that were thought to influence disease prognosis such as size and extension of the primary tumor (T), presence and extent of regional lymph node involvement (N) and presence of distant metastases (M). In 1977, the American Joint Committee on Cancer (AJCC) published their first breast cancer staging system report, which was based on the TNM staging system. Following, in 1985, the International Union Against Cancer presented a clinical classification for breast cancer based on the TNM classification system.[23]

2.3.2 Overview of the TNM Staging Principles

The TNM staging system encompasses four different classifications: clinical (*cTNM*), pathologic (*pTNM*), recurrence (*rTNM*) and autopsy (*aTNM*).[23]

cTNM: Is utilized in order to make local/regional treatment recommendations. It is solely based on evidence gathered prior to initial treatment of the primary tumor:

Physical examination, imaging studies (such as mammography and ultrasound) and pathologic examination of the breast or other tissues obtained from biopsy.[23]

Furthermore, additional descriptors when identifying special cases of *cTNM* and *pTNM* classification include: The "m" prefix in cases with multiple tumors and the "y" prefix in cases where classification is performed during or following initial multimodality therapy (for example, neoadjuvant chemotherapy, radiation therapy or both).[23]

pTNM: Pathology is used to assess the extent of disease and to make recommendations for treatment. It incorporates results of clinical staging with additional evidence obtained from surgery and pathologic examinations of the primary tumor, lymph nodes and primary metastases (only if metastases occurred).[23]

rTNM: Is utilized when additional treatment is required for a tumor that has recurred after a disease-free period.[23]

aTNM: Is used for cancers discovered after the death of a patient.[23]

Table 1. TNM Classification for Breast Cancer from the AJCC Cancer Staging Manual, 6th Edition

(Adapted from Singletary et al., 2006)[23]

Classification	Definition
Primary tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
Tis (DCIS)	Ductal carcinoma in situ
Tis (LCIS)	Lobular carcinoma in situ
Tis (Paget)	Paget disease of the nipple with no tumor (Paget disease associated with a tumor is classified according to the size of the tumor.)
T1	Tumor ≤ 2 cm in greatest dimension
T1mic	Microinvasion ≤ 0.1 cm in greatest dimension
T1a	Tumor >0.1 cm but ≤0.5 cm in greatest dimension
T1b	Tumor >0.5 cm but ≤1 cm in greatest dimension
T1c	Tumor >1 cm but ≤2 cm in greatest dimension
T2	Tumor >2 cm but ≤5 cm in greatest dimension
T3	Tumor >5 cm in greatest dimension
T4	Tumor of any size with direct extension to chest wall or skin, only as described below
T4a	Extension to chest wall, not including pectoralis muscle
T4b	Edema (including peau d'orange) or ulceration of the skin of the breast, or satellite skin nodules confined to the same breast
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma
Regional lymph nodes (N)	
NX	Regional lymph nodes cannot be assessed (eg, previously removed)
N0	No regional lymph node metastasis
N1	Metastasis in movable ipsilateral axillary lymph node(s)
N2	Metastases in ipsilateral axillary lymph nodes fixed or matted, or in clinically apparent* ipsilateral internal mammary nodes in the absence of clinically evident axillary lymph-node metastasis
N2a	Metastasis in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastasis only in clinically apparent* ipsilateral internal mammary nodes and in the absence of clinically evident axillary axillary lymph-node metastasis
N3	Metastasis in ipsilateral infraclavicular lymph node(s), or in clinically apparent* ipsilateral internal mammary lymph node(s) and in the presence of clinically evident axillary lymph-node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph-node involvement
N3a	Metastasis in ipsilateral infraclavicular lymph node(s) and axillary lymph node(s)
N3b	Metastasis in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastasis in ipsilateral supraclavicular lymph node(s)
Regional lymph nodes (pN)†	
pNX	Regional lymph nodes cannot be assessed (eg, previously removed or not removed for pathologic study)
pN0	No regional lymph node metastasis histologically, no additional examination for isolated tumor cells‡
pN0(i-)	No regional lymph node metastasis histologically, negative immunohistochemical staining
pN0(i+)	Isolated tumor cells identified histologically or by positive immunohistochemical staining, no cluster >0.2 mm§
pN0(mol-)	No regional lymph-node metastasis histologically, negative molecular findings (RT-PCR)¶
pN0(mol+)	No regional lymph-node metastasis histologically, positive molecular findings (RT-PCR)¶
pN1	Metastasis in one to three axillary lymph nodes, and/or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent*
pN1mi	Micrometastasis (>0.2 mm, none >2.0 mm)
pN1a	Metastasis in one to three axillary lymph nodes
pN1b	Metastasis in internal mammary nodes with microscopic disease detected by sentinel lymph-node dissection but not clinically apparent*
pN1c	Metastasis in one to three axillary lymph nodes** and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph-node dissection but not clinically apparent*
pN2	Metastasis in four to nine axillary lymph nodes, or in clinically apparent* internal mammary lymph nodes in the absence of axillary lymph-node metastasis
pN2a	Metastasis in four to nine axillary lymph nodes (at least one tumor deposit >2.0 mm)
pN2b	Metastasis in clinically apparent* internal mammary lymph nodes in the absence of axillary lymph-node metastasis
pN3	Metastasis in 10 or more axillary lymph nodes, or in infraclavicular lymph nodes, or in clinically apparent* ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes with clinically negative microscopic metastasis in internal mammary lymph nodes; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastasis in 10 or more axillary lymph nodes (at least one tumor deposit >2.0 mm), or metastasis to the infraclavicular lymph nodes
pN3b	Metastasis in clinically apparent* ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph-node dissection but not clinically apparent*
pN3c	Metastasis in ipsilateral supraclavicular lymph nodes
Distant metastasis (M)	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis

2.3.3 Histopathologic Grading

Aside from stage, histopathologic grading is associated with overall and metastasis free survival (defined from the date of curative treatment to the date of diagnosis of distant metastasis). Also, histopathologic grading is helpful for choosing the appropriate therapy. Knowledge of the tumor grade avoids under treatment of high grade tumors (grade 3) and over treatment of low grade tumors (grade 1). In comparing high and low grade tumors, high grade tumors are considered aggressive.[24] To date, the Scarff-Bloom-Richardson (SBR) tumor grading system has been the most widely accepted and utilized tumor grading system in clinical settings. The SBR tumor grading system examines three grade features: (1) Ductoglandular formation (percentage of cancer composed of tubular structures), (2) Nuclear pleomorphism (change in cell size and uniformity), and (3) Mitosis count (rate of cell division). In calculating the histopathologic grade score, each above mentioned feature is assigned a score between 1-3 (1 indicating more normal histologic or cellular appearance or slower cell growth and 3 indicating more abnormal histologic or cellular appearance or faster cell growth).[24] Then, the score of each feature is added together for a final score that ranges from 3-9. A tumor with a final score of 3, 4 or 5 is considered a grade 1 tumor. A sum of 6 or 7 is considered a grade 2 tumor and a sum of 8 or 9 is considered a grade 3 tumor.

2.4 Epidemiology of Breast Cancer

2.4.1 Incidence and Mortality Trends: An International Perspective

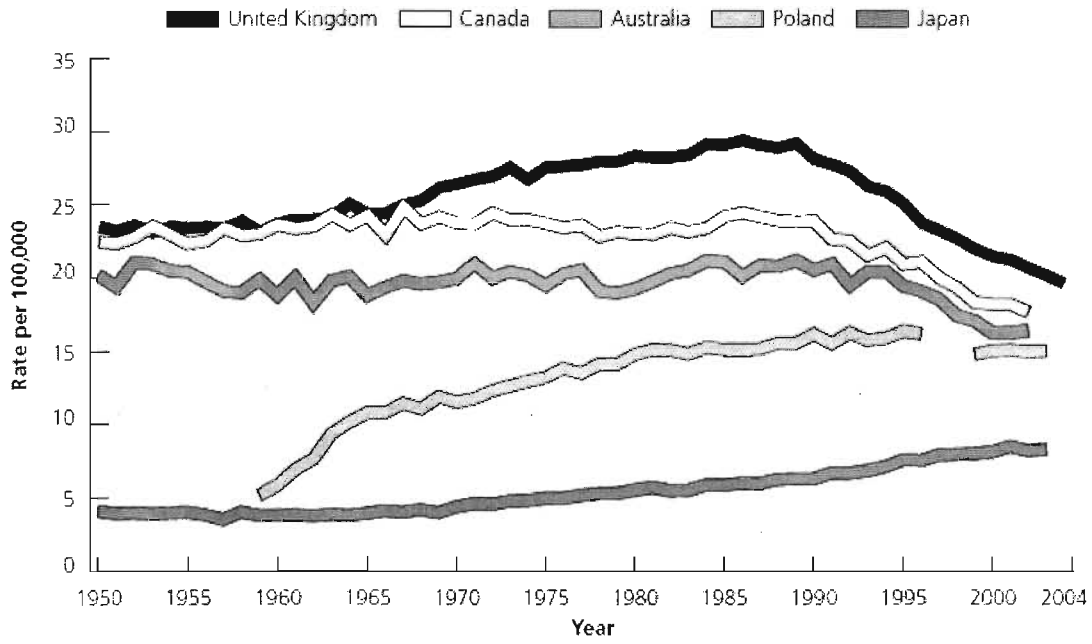
According to the American Cancer Society report on Global Cancer Facts & Figures 2007, breast cancer is the most frequently diagnosed cancer in women. Age-standardized breast cancer incidence rates vary internationally, with the highest number of cases

reported in North America, Australia and Northern and Western Europe. In contrast, Eastern Europe reported intermediate levels whereas large parts of Africa and Asia reported the lowest levels. In 2002, breast cancer incidence rates varied internationally by more than 25-fold.[25] In developing countries the variation in incidence was attributed to low screening rates, lack of cancer registries or incomplete reporting of breast cancers. Age-standardized breast cancer incidence rates ranged from 3.9 per 100,000 in Mozambique to 101.1 per 100,000 in the US. In 2007, approximately 1.3 million new cases of invasive breast cancer and an estimated 465,000 breast cancer deaths were predicted to occur world-wide.[25]

Over the past 25 years the incidence rate for breast cancer increased by 30% in developed countries, resulting from a shift in reproductive patterns and an increase in screening adoption.[25, 26] Additionally, the increase in prevalence of obesity has been associated with an increase in postmenopausal breast cancer risk.[26] Also, since 1981 there has been controversy regarding the role of hormones on breast cancer development. It has been noted that current use of oral contraceptives modestly increases the risk of breast cancer. Furthermore, postmenopausal hormone replacement therapy (HRT) increases breast cancer risk (the risk is exacerbated with prolonged use).[3, 26]

On the other hand, breast cancer mortality rates over the past 25 years have been leveling off or decreasing in most developed countries. As depicted in *Figure 4,* mortality rates in the United Kingdom, Canada and Australia have been decreasing, possibly resulting from improvements in treatment and an increase in mammography screening.[25] On the other hand, trends in mortality rates have not been favourable in Poland and

Japan. It has been speculated that dietary exposure specifically, high fat diets are responsible for the increase in breast cancer incidence and subsequently death.[27]



Note: Break in trend indicates missing data.

Figure 4. **Trends in Age-Standardized Death Rates of Breast Cancer in Five Countries.**[25]

2.4.2 Incidence and Mortality Trends in US Women

Breast cancer accounts for approximately 1 in 4 cancers diagnosed among US women.[3] According to the Breast Cancer Facts & Figures 2009-2010 report put forth by the American Cancer Society, breast cancer incidence tends to increase with age. During the period of 2002-2006 women between the ages of 20-24 reported the lowest incidence rate (1.4 per 100,000) whereas women aged 75-79 reported the highest incidence rate (441.9 per 100,000).[3] However, breast cancer incidence after 80 years of age decreased possibly due to a reduction in screening.[3]

Race. As depicted in *Figure 5*, White women tend to have a higher breast cancer incidence rate beginning at age 45 years. On the other hand, Blacks are at higher risk before the age of 45 years and are more likely to die from breast cancer at every age.[3]

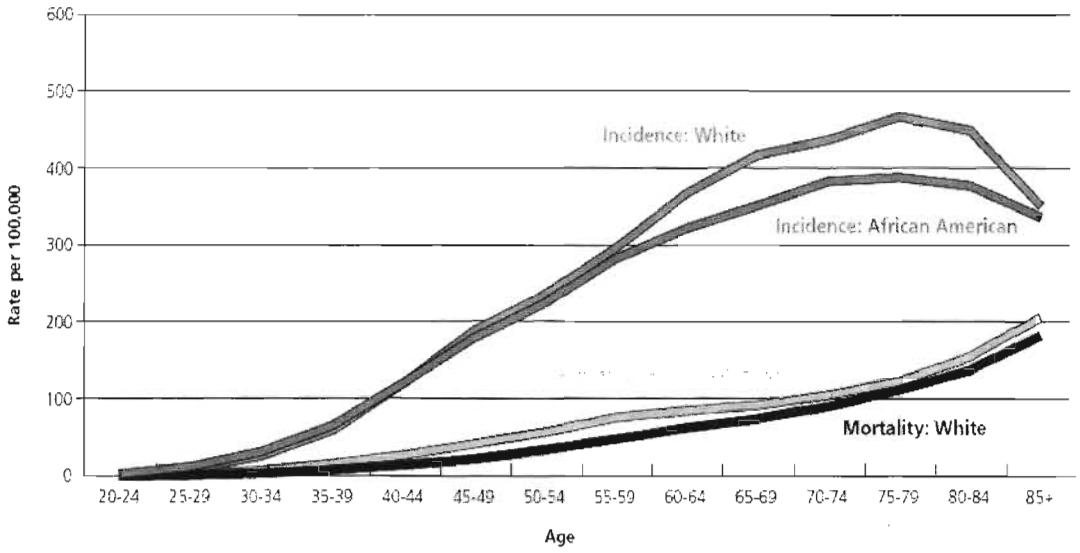


Figure 5. **Female Breast Cancer– Incidence and Mortality Rates by Age and Race, US, 2002-2006.**[3]

Calendar year. Incidence rates have been inconsistent for invasive breast cancer during 1975-2006. During 1975-1985, the incidence rate was constant whereas after 1980 the incidence rate rose by 4.0% annually until 1987. Then, a constant pattern was observed during 1987-1994, rising again by 1.6% per year during the period of 1994-1999. Lastly, during 1999-2006, the incidence rate decreased by 2.0% per year. As previously mentioned, changes in reproductive patterns and greater utilization of breast cancer screening may have resulted in the increase of breast cancer incidence.[3] However, the decrease in breast cancer incidence observed more recently may have been a result of decreased mammography screening and HRT.[3, 28-30] On the other hand, the incidence rate for in situ breast cancer has been increasing during 1980-1990, due to higher utilization of mammography screening, as observed in *Figure 6.*[3]

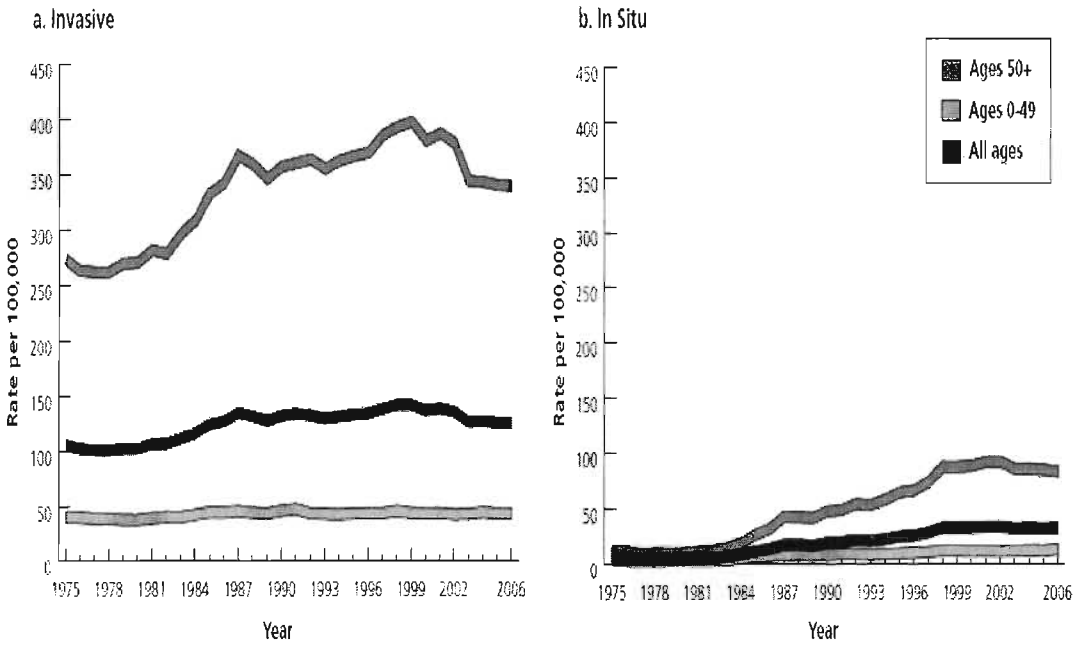


Figure 6. **Incidence Rates of Invasive and In Situ Female Breast Cancer by Age, Adjusted for Delayed Reporting, US, 1975-2006.**[3]

Finally, death from breast cancer has been decreasing since the 1990's.[3, 30] Particularly, from 1990-2006, death rates decreased by 3.2% per year among women less than 50 years and by 2.0% annually in women 50 years and older.[3, 30, 31]

2.4.3 Incidence and Mortality Trends in Canada

As depicted in *Figure 7*, during 1980-2004, breast cancer incidence and mortality trends have been shifting in Ontario.[32] For example, women between the ages of 50-69 years experienced an increase in incidence across the 1980's and thereafter incidence continued to rise steadily. A similar pattern was observed for women of all ages during 1980-2004. The pattern observed in the 1990's resulted from the consistent use of mammography screening.[32]

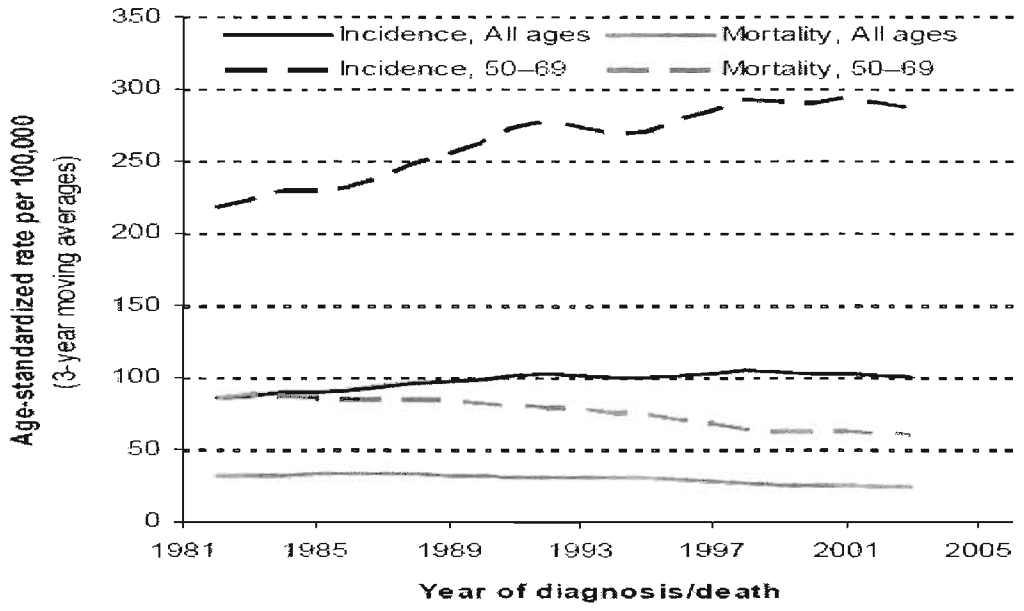


Figure 7. **Breast Cancer Incidence and Mortality Rates in Ontario, 1981-2004.**[32]

Between 1989-2004, death from breast cancer decreased by 33%.[32] The decrease in breast cancer-specific death was attributed to improvements in treatment and screening.

In summary, comparable patterns for breast cancer incidence and mortality were observed in American and Canadian women, for the most part an increase in incidence and a decrease in mortality. Generally in major westernized countries, the incidence has been increasing while mortality has leveled off or decreased.

2.5 Breast Cancer Risk Factors

Modifiable and non-modifiable breast cancer risk factors play a key role in breast carcinogenesis. Modifiable risk factors are preventable whereas non-modifiable risk factors cannot be prevented. Modifiable risk factors include: being overweight or obese after menopause, use of HRT, excess alcohol consumption, physical inactivity and age at first full-term pregnancy.[3, 30] On the other hand, non-modifiable breast cancer risk factors

include: age, family history, benign breast disease, inherited genetic mutations in breast cancer susceptibility genes BRCA1 and BRCA2, early menarche, age at menopause and menopausal status, breast density and atypical hyperplasia.[3, 30, 33, 34] Women encompassing any of the aforementioned risk factors are considered to be at higher risk for breast cancer. *Table 2* outlines the previously mentioned risk factors and additional breast cancer risk factors in order of strength of their association.

Table 2. **Factors that Increase the Relative Risk for Breast Cancer in Women**
(Adapted from Breast Cancer Facts & Figures, 2009-2010)[3]

Relative Risk	Factor
>4.0	Female Age (65+ vs. <65 years, although risk increases across all ages until age 80) Certain inherited genetic mutations for breast cancer (BRCA1 and/or BRCA2) Two or more first-degree relatives with breast cancer diagnosed at an early age Personal history of breast cancer High breast tissue density Biopsy-confirmed atypical hyperplasia
2.1-4.0	One first-degree relative with breast cancer High-dose radiation to chest High bone density (postmenopausal)
1.1-2.0	
Factors that affect circulating hormones	Late age at first full-term pregnancy (>30 years) Early menarche (<12 years) Late menopause (>55 years) No full-term pregnancies Never breastfed a child Recent oral contraceptive use Recent and long-term use of estrogen and progestin Obesity (postmenopausal)
Other factors	Personal history of endometrial or ovarian cancer Alcohol consumption Height (tall) High socioeconomic status Ashkenazi Jewish heritage

Occupational risk factors: Exposure to toxins and light at night has been under investigation and are considered risk factors for breast cancer. Steenland and colleagues found that increased exposure to ethylene oxide, a fumigant used to sterilize surgical instruments, was shown to cause breast cancer in animals. Also, ethylene oxide was associated with higher breast cancer risk among women employed in commercial sterilization facilities.[35] Additionally, flight attendants who experience circadian rhythm disruption by crossing multiple time zones are at an increased risk for breast cancer.[3] Similar studies conducted on night shift work and exposure to light found concordant results. In flight attendants and nurses it was noted that the increase in breast cancer risk was a result of low melatonin levels that occur because of exposure to light at night.[3]

2.5.1 The Role of Menopausal Status and BMI on Breast Cancer Risk

It is suggested that BMI is inversely related to breast cancer risk among premenopausal women whereas postmenopausal women experience a weak to moderate increase in breast cancer risk with increasing BMI.[13, 36, 37] However, results have been inconsistent. Majority of case-control studies found that BMI was positively associated with postmenopausal breast cancer whereas cohort studies demonstrated weak or null associations.[36]

Trentham-Dietz and colleagues used a case-control study to compare incident cases of breast cancer to controls in order to determine the relationship between BMI and breast cancer risk.[7] BMI was reported in quintiles and logistic regression was used to adjust for parity, age at first birth, age at menarche, family history of breast cancer, recent alcohol consumption, education, and for postmenopausal women, age at menopause. In analysis, BMI was weakly associated with decreased risk of premenopausal breast cancer. In multivariate logistic regression, for every 1 kg/m^2 increase in BMI there was a 2% decrease in premenopausal breast cancer risk (OR= 0.98; 95% CI= 0.97-1.00; p= 0.03).[7] On the other hand, high BMI was associated with increased risk of postmenopausal breast cancer. Moreover, adjusted for the aforementioned covariates, per 1 kg/m^2 increase in BMI there was a 3% increase in postmenopausal breast cancer risk (OR= 1.03; 95% CI= 1.02-1.04; $p < 0.001$).[7]

Huang and colleagues prospectively examined BMI at 18 years and at midlife along with adult weight change in relation to subsequent risks of premenopausal and postmenopausal breast cancer incidence and mortality among a large cohort of US female nurses (the Nurse's Health Study cohort). The age-adjusted relative risk (RR) for breast

cancer was close to the null for current BMI up to 26 kg/m^2 however declined with greater adiposity in premenopausal women.[36] In contrast, current BMI and breast cancer incidence in postmenopausal women was positively associated. However, this association was not significant. In addition, stratified analyses by hormone use demonstrated a stronger positive association between current BMI and breast cancer risk in those women who never used HRT (RR= 1.59; 95% CI= 1.09-2.32; $p < 0.001$).[36]

2.5.2 Etiologic Mechanisms in Obese Premenopausal Breast Cancer Patients

The difference in premenopausal and postmenopausal breast cancer risk is that obese premenopausal women on average have longer, more irregular menstrual cycle lengths and a greater tendency for anovulatory cycles.[4, 7, 38-41] The increase in anovulation is a result of low serum estradiol and progesterone levels on target breast cells and a consequent reduced risk of breast cancer.[4, 42]

2.5.3 Etiologic Mechanisms in Obese Postmenopausal Breast Cancer Patients

In postmenopausal women, endogenously produced estrogens play a key role in breast cancer risk and mortality.[43-45] Estrogens, derived from androgenic steroids in adipose tissue, are more likely to increase breast cancer risk in obese postmenopausal women.[7-10, 36, 42, 45, 46] According to *Figure 8*, the elevation in estrone is mediated by the aromatization, in adipose tissue of the C19 steroid androstenedione, which is secreted by the adrenal glands and postmenopausal ovaries. As a result, levels of androstenedione and aromatase are high in obese postmenopausal women. Then, adipose tissue derived estrone is available for conversion to the more biologically potent estradiol. Finally, the circulating levels of estrone and estradiol are positively correlated with body weight.

Elevation in these estrogens has been associated with increased breast cancer risk in obese postmenopausal women.[18, 40, 47]

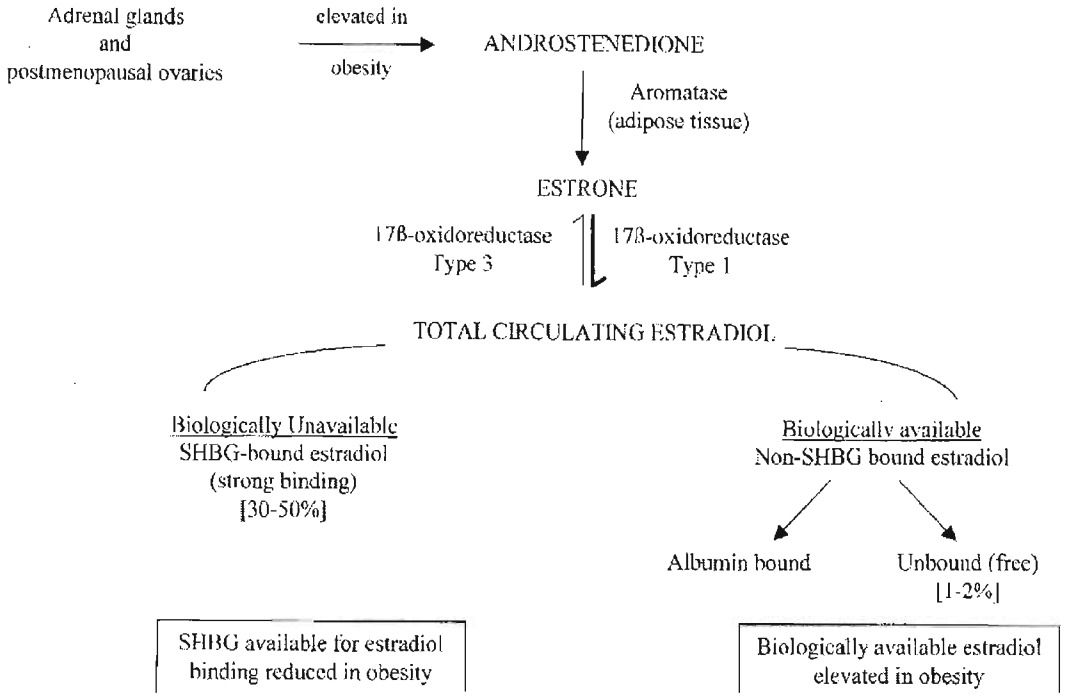


Figure 8. **The Metabolic Production of Estrone and Estradiol from the C19 Steroid Androstenedione and the Bioavailability of Estradiol in Postmenopausal Women.**[40]

However, this relationship diminishes in postmenopausal women who take exogenous estrogen from HRT.[48] HRT masks the aforementioned association because exogenous hormones artificially elevate the amount of circulating estrogens to such a degree in lean and obese women; that the increase in circulating estrogens in adipose tissue is negligible.

In addition, obese women have lower levels of sex hormone-binding globulin (SHBG) which may increase breast cancer risk by raising serum free estradiol and free testosterone levels.[7, 9, 36, 41, 46, 47, 49, 50] Thus, excess estrogen and testosterone may influence both the characteristics and growth rate of the breast tumor.[51]

McTiernan and colleagues examined the association between adiposity and concentrations of androgens, estrogens and SHBG in a population-based, multicenter, multiethnic prospective cohort of 1,185 breast cancer patients.[44] Of the 1,185 breast cancer survivors, 503 were postmenopausal. Obese women (BMI > 30 kg/m^2) had a 35% higher concentration of estrone compared to women with a BMI < 22 kg/m^2 (p= 0.005).[44] Moreover, estradiol concentrations increased by 130% in obese women when compared to their lighter counterparts (p= 0.002).[44] Additionally, concentrations of testosterone increased as adiposity increased and as a result, obese women had twice the testosterone concentration when compared to their lighter counterparts (p= 0.0001).[44] Furthermore, free estradiol and testosterone levels were significantly higher in overweight and obese women compared to the lightest women (p= 0.0001). Finally, SHBG levels significantly decreased with increasing BMI (p= 0.0001).[44]

2.5.4 Insulin and Insulin-Like Growth Factors (IGFs), Menopausal Status and Breast Cancer Risk in Obese Women

According to *Figure 9,* insulin and IGF are found in high levels in obese women and have shown to express mitogenic effects on normal and neoplastic breast epithelial cells; particularly in premenopausal women because insulin and IGF potentiate the effects of circulating estrogens.[20, 21, 41, 45, 46] Hyperinsulinemia and insulin resistance syndrome have been involved in the mechanisms by which excess weight increases premenopausal breast cancer risk. Furthermore, insulin and IGF exert their effect by reducing circulating levels of SHBG and as a result increasing circulating estrone and estradiol levels. For example, Del Giudice and colleagues demonstrated that elevated plasma insulin concentrations

were associated with premenopausal breast cancer risk after adjusting for age and body weight.[54]

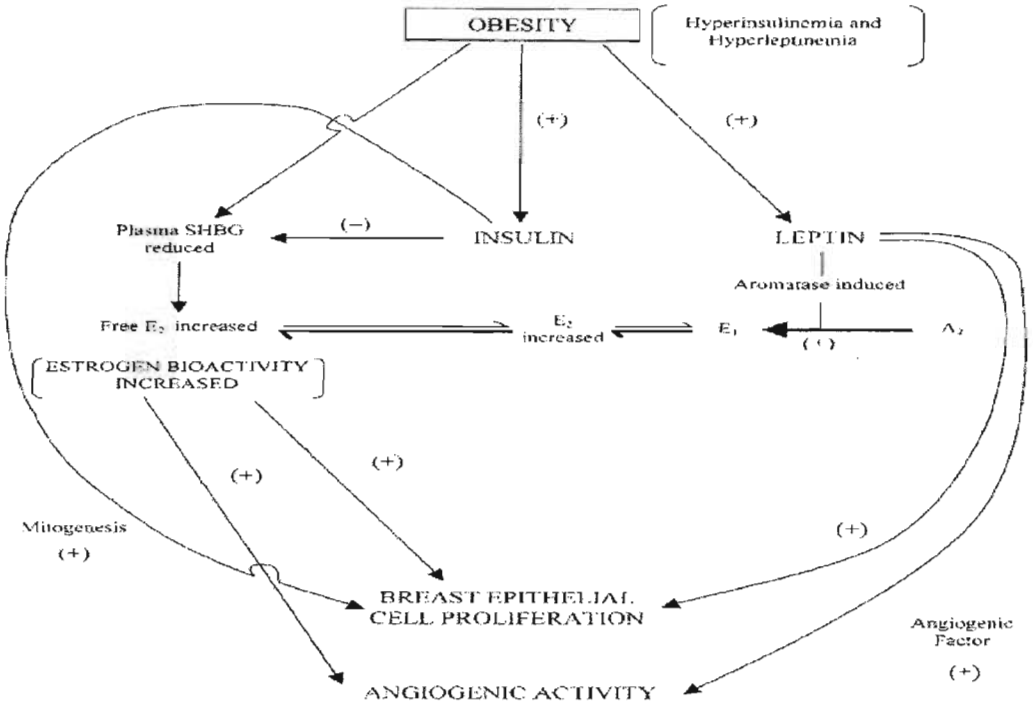


Figure 9. **Obesity and the Interaction between Estradiol and Sex Hormone-Binding Globulin (SHBG) and Insulin and Leptin in Breast Cacner Cell Proliferation and Angiogenesis.**[40]

Also, elevated plasma IGF-1 levels were associated with an increased risk of breast cancer. However, this relationship was only observed in premenopausal women.[40] For example, Li and colleagues extracted total and free plasma IGF-1 and IGF-binding protein-3 (IGFBP) concentrations in 40 newly diagnosed premenopausal and postmenopausal breast cancer patients and 40 race and age matched healthy controls.[55] In analysis, there was no significant difference in mean BMI between cases and controls. Adjusted for menopausal status, free IGF-1 but not total IGF-1 was significantly

associated with an increased breast cancer risk. Also, a high IGF-1/IGFBP-3 ratio was a significant risk factor for breast cancer development.[55]

2.5.5 Mechanisms of Leptin and Breast Cancer Risk in Obese Women

Leptin is a neuroendocrine hormone solely produced in adipose tissue. It is suggested that the biological actions of leptin are associated with obesity and breast cancer development and progression.[8, 56] Leptin fuels the growth of human breast cancer cell lines in vitro and the expression of proteolytic enzymes, which are necessary for breast tissue invasion.[40, 57-59]

Like insulin, leptin can modify estrogenic activity and therefore, has the ability to influence the biological behavior of estrogen-dependent breast cancers. The mechanism of action is by way of estrogen production because leptin can induce the activity of aromatase (an enzyme responsible for the biosynthesis of estrone from androstenedione in adipose tissue).[40, 45] Tessitore and colleagues conducted a case-control study and found elevated plasma leptin levels and an increase in adipose tissue leptin mRNA levels in 23 breast cancer patients.[60] However, the number of cases and controls were too few to draw appropriate conclusions.

2.6 Breast Cancer Prognosis

2.6.1 Defining Breast Cancer Mortality and Survival

Mortality or mortality rate and survival or survival rate are often incorrectly used interchangeably. Mortality rate is the number of deaths per population (usually 100,000) per unit time whereas survival measures the time it takes to develop an outcome of interest. Survival can be measured using Cox proportional hazards regression which uses time-to-event survival analysis. Survival analysis examines and models the time it takes

for an event to occur. In this study, survival was defined from the time of primary breast cancer diagnosis to time of breast cancer-specific death. In studying survival, multiple outcomes can be assessed in patients diagnosed with breast cancer. The four distinct outcomes are: ***breast cancer recurrence/progression, breast cancer-specific survival, competing-causes survival and all cause or overall survival.***[61, 62]

1) ***Breast cancer recurrence/progression:*** (a) ***Recurrence.*** Local/regional cancer recurrence following a curative resection, and (b) ***Progression.*** Local spread or distant recurrence/metastasis following an attempted cure.
2) ***Breast cancer-specific survival:*** Time from primary breast cancer diagnosis to time of breast cancer-specific death.
3) ***All cause or overall survival:*** Time from primary breast cancer diagnosis to time of death from any cause.
4) ***Competing-causes survival:*** Time from primary breast cancer diagnosis to time of death from any cause other than breast cancer.

This study will examine breast cancer recurrence/progression and breast cancer-specific death.

2.6.2 Five-Year Relative Survival Ratio (RSR): Canadian Population

Like mortality, survival is also an important indicator of cancer burden. Survival is a measure of disease severity. Survival estimates are used to establish priority areas for improving cancer prognosis and control.[63] Population-based survival estimates are expressed as ratios. RSR is defined as the ratio of the observed survival for a group of people diagnosed with cancer to the survival expected for people of the same general population.[64]

Between the period of 2002-2004, the five-year RSR for breast cancer in Canadian women was 87% (95% CI= 87-88%). In interpreting the RSR, those women diagnosed with breast cancer between 2002-2004 were estimated to be 87% as likely to live for another five-years when compared to members of the general population.[63]

RSR by province, 2002-2004. The age-standardized breast cancer five-year RSR for males and females combined in Canada was 87% (95% CI= 87-88%). For example, Ontario (88%; 95% CI= 87-88%) and Alberta (88%; 95% CI= 86-89%) had the highest five-year RSR whereas Prince Edward Island (84%; 95% CI= 79-89%) and Nova Scotia (84%; 95% CI= 82-86%) reported the lowest five-year RSR rates.[63] A possible explanation for the difference in the five-year RSR among provinces is that provincial regulations for screening and early detection may be different for each province. In addition, access to specialized cancer treatment and differences in population characteristics may also contribute to the different five-year RSR's observed by province.

RSR by age, 2002-2004. The five-year RSR for breast cancer was different for all age groups. However, the five-year RSR was lowest among those who were diagnosed at a later age. For instance, women diagnosed with breast cancer between the ages of 60-69 had the highest five-year RSR, 90% (95% CI= 89-90%). On the other hand, persons diagnosed between 80-99 years had the lowest five-year RSR, 80% (95% CI= 77-82%).[63]

Generally, survival tends to decrease with age because individuals diagnosed with cancer receive less therapy due to the presence of comorbidities. The presence of comorbidities reduces the body's ability to tolerate and respond to treatment.[63] This pattern is evident in those women diagnosed with breast cancer as the five-year RSR by

age decreased in women between 60-69 years old vs. 80-99 years old, from 90% to 80%.[63]

RSR estimates for 2002-2004 vs.1992-1994. The five-year RSR for breast cancer increased approximately 5% between 1992-1994 and 2002-2004.[63]

2.6.3 Five-Year Survival Rates: US Population

According to Surveillance, Epidemiology, and End Results (SEER), the overall five-year relative survival during 1999-2006 from 17 SEER geographic areas was 89.0%. Furthermore, the relative five-year survival by race was: 90.2% for White women and 77.5% for Black women.[65]

The relative five-year survival rate by stage at diagnosis for all races combined was: 98.0% (localized), 83.6% (regional), 23.4% (distant) and 57.9% (unknown), respectively.[65]

2.6.4 Covariates Predicting Breast Cancer Survival

In exploring previous literature, Shek and Godolphin derived a multivariate Cox model that incorporated prognostic factors that demonstrated independent predictive value for overall survival.[66] Clinicopathologic variables collected at time of breast cancer diagnosis were: stage, based on the TNM staging classification system, axillary node status, ER status, age and menopausal status. Stepwise regression was utilized to identify important covariates, at a chosen significance level of $p \leq 0.05$.

Using stepwise regression, nodal status was the primary variable forced into Cox models as it produced the maximum likelihood. Followed by stage (as stage improved the likelihood of the preceding variable), demonstrating that individuals who were in a higher stage of their disease had more detrimental effects. Then, the negative coefficient of log

ER was added and was indicative of a reduced risk associated with increasing ER concentration. Also, tumor necrosis was associated with lower overall survival when compared to minimal or no necrosis.[66] All covariates were added in the same manner and decision not to include additional covariates to Cox models was based on the above mentioned significance level.

The study conducted by Shek and Godolphin contained strengths and limitations. The study had a large sample size (n= 1,184) and a follow-up of 10 years. Also, an extensive set of prognostic variables were available for analysis. However, the study did possess limitations. A major flaw was the criteria used for prognostic covariate selection. Stepwise regression and a significance level of $p \leq 0.05$ for developing prediction models is not a highly regarded technique.[67] Instead, using *a priori* covariate selection alongside "backward" elimination, with a level of significance of $p \leq 0.2$ (however this is not a fixed value) for predictor selection is the preferred method because it provides better prediction.[67]

Newman and colleagues examined whether obesity was related to breast cancer-specific survival after covariate adjustment.[10] The following variables were tested in univariate Cox models and deemed significant predictors of breast cancer-specific survival: age (p= 0.03), size of tumor, number of positive nodes, clinical stage, ER/PR status (each with $p < 0.001$), chemotherapy and radiation therapy (p= 0.02).[10] Furthermore, tumor size, number of positive nodes and ER status were highly statistically significant in multivariate models adjusted for any clinical and histopathologic predictors. Also, adjusted for tumor size, number of positive nodes and ER status; PR status, stage and radiation therapy ceased to be statistically significant. Chemotherapy maintained a

marginal level of statistical significance.[10] The study conducted by Newman and colleagues found concordant results with Shek and Godolphin, that tumor size, node positive tumors and low ER status correlates with poor breast cancer-specific survival and overall survival.[10, 66] Some studies also found that radiation therapy, HRT and tamoxifen were important predictors of overall survival.[4, 9, 68] Additional prognostic factors associated with breast cancer survival included: race, history of benign breast disease, education level, menopausal status, oral contraceptive use, age at menopause, protein intake, physical activity and smoking.[4, 9, 68] However, there are issues surrounding the above mentioned prognostic variables. Many of these prognostic variables are only associated with survival in univariate but not multivariate analysis. Also, some of the aforementioned prognostic variables are only associated with competing causes survival and not breast cancer-specific survival. For this study, well established prognostic factors chosen *a priori* and from past literature will include: age, stage, ER/PR status, surgery, radiation therapy and chemotherapy.[10, 66, 69-71]

2.6.5 Mechanisms Underlying Low BMI and Breast Cancer Prognosis

Moon and colleagues suggest that a mechanistic relationship exists between low BMI and poor breast cancer outcomes after treatment. This process involves tumor-modulating roles of local and systemic adipocytes. Animal experiments have demonstrated that the presence of mammary adipocytes is critical for mammary gland development and irradiation of mammary fat pads caused malignant progression of normal mammary cells.[61] To date, the above mentioned mechanistic relationship is not fully understood and future research should further examine this relationship.

2.7 Obesity and Breast Cancer Survival

2.7.1 Obesity and Overall Survival

Obesity is one prime example of a modifiable risk factor that can be controlled through diet and physical activity.[3, 72, 73] Even though obesity is a modifiable risk factor, the prevalence of obesity in women has nearly doubled during the past decade and the increase in prevalence is evident globally, in the US and Canada.[4-6, 74] To date, literature suggests that women who are overweight or obese at breast cancer diagnosis are found to have worse overall and recurrence-free survival.[8]

2.7.2 Measuring Obesity

Obesity is defined as the buildup of adipose tissue in excess and to an extent that impairs both physical and psychosocial health and well-being.[75] Even though it is possible to measure body fat directly, these methods are expensive, time-consuming and do not pertain to fieldwork and clinical practice. Instead, the Quetelet Index (QI) or BMI is used as a proxy to measure body fat. BMI is calculated using weight in kilograms (kg) divided by height in meters squared (m^2).[74, 75] BMI is a standardized classification system used to compare obesity prevalence rates worldwide. Furthermore, BMI categories are used to develop plans for health management and to track changes in the obesity epidemic. A widely utilized classification system is the World Health Organization (WHO) international classification system of adult underweight, normal, overweight and obesity, as outlined in *Table 3*.[5, 75-78]

Table 3. **The International Classification System of Adult Underweight, Normal, Overweight and Obesity According to BMI**
(Adapted from WHO, 2009)[78]

Classification	BMI (kg/m^2) Principal cut-off points
Underweight	<18.50
Severe thinness	<16.00
Moderate thinness	16.00-16.99
Mild thinness	17.00-18.49
Normal range	18.50-24.99
Overweight	≥25.00
Pre-obese	25.00-29.99
Obese	≥30.00
Obese class I	30.00-34.99
Obese class II	35.00-39.99
Obese class III	≥40.00

2.7.3 Using BMI as a Proxy to Measure Body Fat Percent

Liu and colleagues measured the consistency between BMI and body fat percent in 200 Taiwanese women diagnosed with breast cancer.[79] Body fat and fat-free mass were measured by bioelectrical impedance one day before breast surgery and BMI was calculated using weight in kilograms (kg) divided by height in meters squared (m^2). The relationship between BMI and fat/lean mass was tested by Pearson's correlation coefficient (r). Furthermore, accuracy was calculated using sensitivity and specificity to reflect the diagnostic performance of BMI in detecting body fat percent defined obesity. In analysis, BMI was strongly correlated with fat mass ($r= 0.96$; $p < 0.001$) and body fat percent ($r= 0.91$; $p < 0.001$). In addition, the correlation between BMI and fat-free mass was moderate and significant ($r= 0.49$; $p < 0.001$). However, BMI demonstrated poor sensitivity for identifying obesity (47%; 95% CI= 39-55) and excellent specificity (99%; 95% CI= 92-100).[79] This study contained limitations such as a small sample size and

differences in ethnic variation. The study sample was limited to Asians thus results were not externally generalizable as different BMI cut-points may need to be utilized for different races. For instance, in defining obesity using BMI in Black women there is a potentially confounding influence because of their high bone mineral density and skeletal muscle mass; resulting in an elevated BMI. In order for BMI to be an accurate measure of adiposity, BMI requires that body weight is a true depiction of adiposity and that adult height is not influenced by extraneous pathologic conditions. Despite BMI's limitation as a surrogate measure of body fat, to date, BMI is considered convenient and simple in clinical practice and research.

2.7.4 Obesity and Breast Cancer Recurrence/Progression

Loi and colleagues examined the effect of obesity on distant recurrence in newly diagnosed breast cancer patients. Time to failure was considered from the date of breast cancer diagnosis. Date of distant recurrence (abstracted from medical records) was used as the time of failure. Women who were not known to have had a distant recurrence but died were assumed to have failed at the date of death. In analysis, 1,101 cases were used for time to distant recurrence.[80] Eight hundred-thirteen women were premenopausal and 288 women were postmenopausal. During follow-up, 264 distant recurrences were observed. In unadjusted analysis, obesity (BMI > 30 kg/m^2) was associated with an increased risk of distant recurrence from breast cancer (HR= 1.50; 95% CI= 1.07-2.09; p= 0.02) .[80] Also, age, tumor grade, nodal status and PR hormone positive status were found to independently predict distant recurrence. Since majority of patients were premenopausal, estrogen levels influenced by peripheral aromatization in adipose tissue is insignificant compared with that produced by functioning ovaries. Additionally, in

obese premenopausal women, the hormonal make-up is different and obesity has been associated with unchanged or decreased levels of estrogen.[80] In multivariate analysis, obesity remained an independent prognostic factor for distant recurrence (HR= 1.57; 95% CI= 1.11-2.22; p= 0.02). Specifically, in premenopausal women the HR was 1.50 (95% CI= 1.00-2.26; p= 0.06) and in postmenopausal women the HR was 2.03 (95% CI= 0.99-4.21; p= 0.07).[80] However, results were not significant but are trending to significance.

On the other hand, Kroenke and colleagues examined the association between BMI and breast cancer recurrence in 5,204 women diagnosed with invasive breast cancer.[9] After breast cancer diagnosis, if a second cancer was reported (in the lungs, liver, bone or brain), it was assumed that breast cancer had recurred. Also, recurrence included those women who did not report a second cancer but who died as a result of breast cancer. It was assumed that disease recurred two or more years before death. Follow-up ranged between 2-26 years, with a median follow-up of nine years. During follow-up, 681 breast cancer recurrences were observed. Multivariate analysis demonstrated a non-significant relationship between BMI prior to breast cancer diagnosis. Moreover, results were not significant for breast cancer recurrence between obese women (BMI > 30 kg/m^2) and non-obese women (BMI of 21-22 kg/m^2).[9]

2.7.5 Obesity and Breast Cancer-Specific Death

Zhang and colleagues prospectively studied 698 women diagnosed with postmenopausal unilateral breast cancer.[81] The study determined whether pre-diagnosis obesity, body fat distribution and dietary intake of fats, antioxidants and fibers were associated with overall survival and breast cancer-specific death. During a six year follow-up, 56 deaths were reported. Of the 56 deaths, 40 had breast cancer among the causes listed on their

death certificates. In analysis, restricted to the 40 breast cancer specific-deaths, the adjusted RR for the highest (28.8-45.9 kg/m^2) vs. lowest (16.0-24.6 kg/m^2) tertile of BMI was RR= 1.50 (95% CI= 0.70-2.90; p= 0.26).[81] Women in the highest tertile of BMI experienced a 50% higher risk of death from breast cancer in comparison to the lowest tertile. However this relationship was not significant. A possible explanation for the non-significant result was the short follow-up and low study power (low number of events).

Daling and colleagues conducted a population-based survival study, including 1,177 premenopausal women diagnosed with invasive ductal breast carcinoma.[82] The outcomes of interest were all cause mortality and breast cancer-specific death. Of the 1,177 women, 317 deaths were observed and of the 317 deaths, 283 were breast cancer-specific. Weight one year prior to diagnosis was used to calculate BMI. In analysis, women in the highest quartile of BMI (25.85-52.57 kg/m^2) had a 1.70-fold increased risk (95% CI= 1.00-2.90; $p < 0.05$) of breast cancer-specific death compared to the lowest quartile of BMI (15.80-20.64 kg/m^2). These results were independent of other prognostic factors such as age, year at diagnosis, lymph node status and tumor characteristics: tumor size, ER/PR status, and alterations in c-*erb* B-2 oncogene protein, apoptosis regulatory protein bcl-2, p53 tumor suppressor gene and cell cycle protein p27.[82]

Whiteman and colleagues also examined the relationship between breast cancer-specific death and factors related to body size such as BMI in adulthood, BMI at age 18 and weight change from 18 years of age to adulthood.[4] During a 14.6 year follow-up, 1,671 women died. Specifically, 1,347 were recorded as deaths due to breast cancer. In analysis, adjusted for age, race, radiation therapy, history of benign breast disease, education, menopausal status and stage, obese women (BMI $\geq$ 30.00 kg/m^2) were

significantly more likely compared to lean women (BMI $\leq$ 22.99 kg/m^2) to die from breast cancer (HR= 1.34; 95% CI= 1.09-1.65).[4] Additionally, women with a BMI of 25.00-29.99 kg/m^2 or 23.00-24.99 kg/m^2 had the highest risk of breast cancer death (HR= 1.25; 95 %CI= 1.08-1.44 and HR= 1.20; 95% CI= 1.04-1.39; p= < 0.0001) compared to women with a BMI of $\leq$ 22.99 kg/m^2.[4] Furthermore, Whiteman and colleagues assessed whether a J-shaped association was present between BMI and breast cancer-specific death. It was observed that women with a BMI of $\leq$ 18.50 kg/m^2 did not have a breast cancer-specific death risk that was significantly different from women with a BMI of 18.51-22.99 kg/m^2 (HR= 1.07; 95% CI = 0.81-1.41). Stratified by menopausal status, the association between high BMI and breast cancer-specific death persisted. Therefore, results between BMI and premenopausal and postmenopausal breast cancer survival are consistent as literature demonstrates a relationship with breast cancer-specific death.[9, 10, 41, 68, 83]

Lastly, Reeves and colleagues determined the effect of BMI on risk of all cause, cardiovascular, competing causes and breast cancer death.[84] Interaction terms were included to evaluate effect modification by age at diagnosis. The study sample comprised of 533 women $\geq$ 65 years of age at study entry and diagnosed with early and late stage breast cancer. During an 8.1 year follow-up, 206 deaths were observed, 45 of which were breast cancer-specific. An obese woman at 65 years of age had a five times higher risk of death attributed to breast cancer compared to a woman with the same age but with a normal BMI (HR= 4.93; 95% CI= 1.12-21.70).[84] However, an obese woman at 85 years of age had a lower risk of breast cancer-specific death compared to a woman of the same age but of a normal BMI (HR= 0.30; 95% CI= 0.08-1.09).[84] This study had limitations.

First, the breast cancer-specific death analysis was underpowered because of the small number of events resulting in wide confidence intervals. The interaction between BMI and age on breast cancer-specific death was borderline statistically significant (p= 0.05) and may have reached statistical significance if more events were observed. Also, the analysis only took into account White females and women were in generally good health at baseline. Thus, these women may not be representative of the general population of breast cancer survivors. Finally, important prognostic data such as treatment was not available therefore the authors were unable to control for treatment in analysis.[84]

2.7.6 Underweightness and Breast Cancer Recurrence and Breast Cancer-Specific Survival

BMI as a prognostic factor remains controversial despite the considerable number of studies. The impact of underweightness (low BMI) on breast cancer recurrence and death has not been adequately addressed because literature has focused primarily at the role of high BMI.[22, 47, 85] However, literature indicates that underweight populations are at increased risk of overall mortality compared to those of normal weight.[86] Thus, it is crucial to understand the prognostic significance of low BMI in breast cancer patients because this may identify a high-risk subgroup.

Moon and colleagues explored the prognostic significance of underweight, overweight and obesity in Korean breast cancer patients.[61] The association between BMI and breast cancer recurrence and death was explored using the Korean Breast Cancer Registry (KBCR) database and the Seoul National University Hospital Breast Cancer Center (SNUHBCC) database. SNUHBCC is a prospectively maintained database providing detailed information on the type and date of breast tumor recurrence and last

follow-up date. All patients were classified according to the WHO international classification system of underweight, normal weight and obese. Multivariate Cox analyses were conducted to study the prognostic significance between BMI and overall survival, breast cancer-specific survival and breast cancer recurrence. Furthermore, stratified Cox regression analysis were performed for overall and breast cancer-specific survival to evaluate the prognostic significance of BMI in each group across stage (modeled using the AJCC criteria; stages I-III).

From the KBCR database, 24,698 patients with non-metastatic, invasive breast cancer were identified for whom information on weight and height was collected at time of initial diagnosis. Overweight patients had larger tumors, were older at diagnosis and had higher frequencies of lymph node metastasis and hormone receptor-negative tumors.[61] In multivariate analysis, overweight and obese patients had non-significant overall and breast cancer-specific survival differences when compared to normal weight patients. On the other hand, low BMI was observed to be an independent negative prognostic factor for overall survival in multivariate analysis. Also, underweight patients had a significantly higher risk of breast cancer-specific survival (HR= 1.49; 95% CI= 1.15-1.93; p= 0.002).[61] Finally, stratifying by stage demonstrated a prognostic effect for underweightness across different disease stages.

Furthermore, to explore the association between BMI and breast cancer recurrence, the SNUHBCC database was used because the KBCR did not include time to recurrence or type of recurrence. From the SNUHBCC database, 4,345 patients with non-metastatic, invasive breast cancer with available BMI data were identified. Underweight patients showed significant differences in disease-free, distant metastasis-free and local

recurrence-free survival. However, the difference was highest for local recurrence (HR= 5.13; 95% CI= 2.66-9.90).[61] In summary, being underweight was an independent prognostic factor for breast cancer survival and for all types of recurrence especially local recurrence. However, overweight and obese patients failed to show a significant difference in breast cancer recurrence and breast cancer survival.

Despite the significant findings between BMI and breast cancer outcomes, the study had limitations. First, the patients in both databases were heterogeneous in terms of disease stage, presence of comorbidities and treatment delivery. Since treatment delivery differs according to body weight status, heterogeneous treatment may confound results.[61] Also, the patients included were all of Asian descent. Thus, results are not generalizable to other ethnic groups due to body size variations between ethnicities. Finally, the nonlinear effect between BMI and breast cancer outcomes was not described. Additional studies need to assess the nonlinear effect of BMI and breast cancer-specific death. This study will describe the nonlinear relationship of BMI and breast cancer recurrence/progression and breast cancer-specific death.

2.8 Nonlinear Relationship between BMI and Breast Cancer Survival and Breast Cancer Recurrence

To date, research has not explored the relationship between BMI and breast cancer-specific survival. However, the two studies that did evaluate the nonlinear relationship of BMI have identified a nonlinear relationship with overall survival and breast cancer recurrence.[71, 87] Studying overall survival may hide any relationships that exist specific to breast cancer. Despite the fact that literature exploring the association between BMI and

overall survival and breast cancer recurrence is available, research should examine this relationship with breast cancer-specific death.

Suissa and colleagues examined the nonlinear relationship between the Quetelet index (QI), overall survival and breast cancer recurrence in 68 women diagnosed with early stage breast cancer.[87] All patients were randomized to receive radical mastectomy, total mastectomy with irradiation or total mastectomy with no further treatment. Patient medical records were reviewed at time of mastectomy. Body weight was measured using the QI, an equivalent measure to the BMI classification system. QI was defined as 0.01 times the weight in pounds divided by the height squared in inches.[87] Cox regression was used to examine the association between the QI, overall survival and breast cancer recurrence, adjusted for age, stage, menopausal status and treatment.

First, the QI was modeled as a linear effect. A non-significant relationship was observed between the QI and overall survival (p= 0.47). However, when the QI was modeled using quadratic terms, a significant association was observed between the QI, overall survival and breast cancer recurrence. Also, the hazard function was concave and indicative of underweight and overweight status being predictive of an unfavorable prognosis of breast cancer. The effect was most evident in those women with a QI of 5 (overweight). However, the hazard of death was lowest in those women with a QI of 3.4 (normal weight).[87]

Despite the significant relationship observed between BMI and overall survival, the study contained limitations and results should be interpreted with caution. A major flaw was the sample size, which consisted of 68 female breast cancer patients. Furthermore, the 68 patients were divided into three treatment groups which would

produce unstable estimates of actual survival probabilities by the Kaplan-Meier product limit method.[87] Instead predicted survival probabilities were reported. Also, there was limited emphasis placed on stage because early stage breast cancer cases were only eligible. Additionally, effect modification was not assessed. This resulted from the small sample size which made it more difficult to detect any interaction effects. Also, the methodology contained limitations. The methodology contained minimal description of the prognostic variables and no explanation was provided as to how and why these variables were included in modeling. Additionally, according to Suissa and colleagues and as depicted in *Figure 10*, the predicted probability of death within five-years of diagnosis was almost 100% for a QI of 5. This is misleading and therefore an inaccurate depiction of the true five-year mortality rate.[87] Finally, study results are in need of validation using a larger sample and higher quality epidemiologic study.

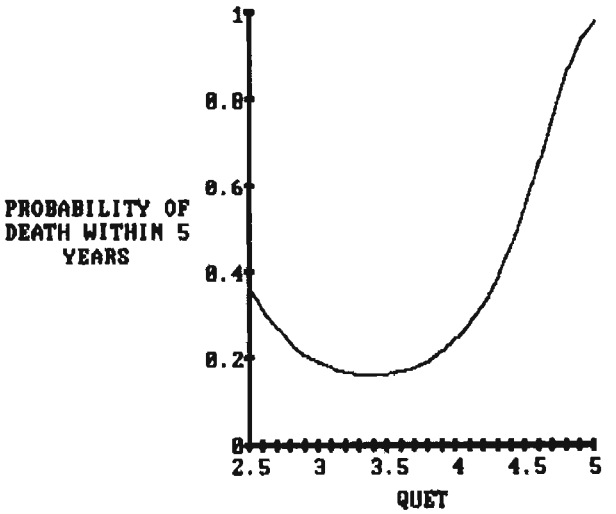


Figure 10. **Predicted Probability of Death within 5 Years of Diagnosis, a Function of QI Obtained from the Cox Regression Model.**[87]

A similar study was conducted by Goodwin and colleagues who sought to confirm the relationship reported by Suissa and colleagues (*Figure 10*).[71, 87] Goodwin and colleagues validated the aforementioned association using a sample of 512 women

diagnosed with early stage breast cancer.[71] The median follow-up was 50 months. During follow-up, 45 women died however all but three deaths were caused by breast cancer. First, the log hazard was modeled linearly. BMI predicted distant disease-free survival (p= 0.047) and marginally predicted overall survival (p= 0.063). Then, the relationship was modeled using quadratic terms. Modeling BMI using quadratic terms demonstrated a significant improvement when compared to the linear model ($p < 0.001$). Also, it identified a significant relationship with distant disease-free and overall survival ($p <$ 0.001). According to Goodwin and colleagues and as depicted in *Figure 11*, women with a high BMI reported the worst outcomes (BMI > 25 kg/m^2). A J-shaped relationship was observed; which was also an identical relationship observed by Suissa and colleagues (*Figure 10*).[71, 87]

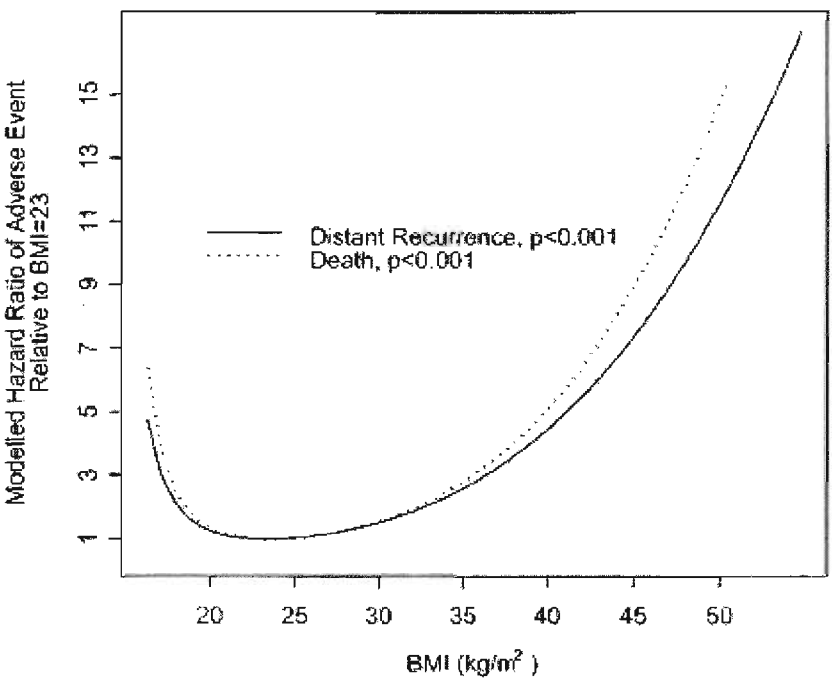


Figure 11. **Hazard Ratio Functions for Distant Recurrence and Death, a Function of BMI Obtained from the Cox Regression Model.**[71]

This study will model BMI using restricted cubic splines as opposed to quadratic terms because quadratic terms are not versatile for three reasons: (1) Quadratic terms cannot describe sudden changes in shape, (2) Quadratic terms can be unstable where the data is thin; usually in extreme ranges such as the tails, and (3) Quadratic terms are not capable of describing a great range of nonlinear patterns.[16, 88, 89] In this study, greater emphasis will be placed on stage, including both early and late stage breast cancer cases. Finally, breast cancer-specific death will be the outcome of interest as opposed to overall survival.

CHAPTER 3: METHODOLOGY

3.0 Overview

The current chapter describes the study design, data collection process and statistical analysis. The methodology section outlines the study design along with a brief description of Henry Ford Health System (HFHS), the place of data collection. Then, the type of data abstracted from patient medical records such as sociodemographic, clinicopathologic and anthropometric data is discussed. Furthermore, an explanation is provided on the outcome data, breast cancer recurrence/progression and breast cancer-specific death. Finally, this chapter ends with a description of the statistical analysis used to address the research questions.

3.1 Study Design and Context

The HFHS dataset was originally designed to study comorbidities and racial disparity among breast cancer patients. However, the dataset contained necessary breast cancer outcome data and relevant anthropometric, clinicopathologic and sociodemographic data. Using the HFHS comorbidity database, a historical cohort study design was implemented to develop a better understanding of the relationship between BMI and breast cancer recurrence/progression and breast cancer-specific death.

Sample size: Patients were diagnosed with primary breast cancer between 1985 and 1990 inclusive, at HFHS. In 2002, 906 breast cancer cases were identified from the HFHS Tumor Registry, an American College of Surgeons, Commission on Cancer–certified registry and were included in the study.

Study context: HFHS is a large, comprehensive, nonprofit system that annually provides medical care to more than 500,000 people. Of those, approximately 30% were

Black. In 1997, the HFHS patient population distribution in 10 age, 2 race and 2 sex categories (40 strata) differed from the metropolitan Detroit (Wayne, Oakland, and Macomb counties, 1990 census data) distribution by 5.3% or less in all strata. These observations suggest that the HFHS patient population is representative of the community it serves.[62]

3.2 Study Criteria

Analysis was limited to Blacks and Whites only. Asians, Pacific Islanders, Native Americans and others were excluded because their numbers were too small to permit meaningful analysis. Four patients were identified as Hispanic and were included in analysis according to the race category they chose (3 White and 1 Black).

3.3 Research Ethics Approval

The HFHS institutional review board approved the study. However, the review board waived the need to obtain patient consent because the study examined patient medical records. The original study was funded by the US Department of Defense (US Army Medical Research and Material Command) grant DAMD17-00-1-0287. Approval to conduct the current secondary data analysis was received from Brock University Research Ethics Board.

3.4 Medical Record Abstraction

3.4.1 Sociodemographic Data

Study data were abstracted from patient medical records. Patient medical records were abstracted dating back to the patient's presenting symptoms or to the initial visit that ultimately led to a breast cancer diagnosis. Sociodemographic data included age, gender, race, marital status and socioeconomic status (SES). SES was estimated using block

group median household income, derived from the patient's address at breast cancer diagnosis and 1990 US census data. Race was classified according to self-report on registration forms.

3.4.2 Clinicopathologic Data

Exposure and clinicopathologic data included stage, treatment and hormone receptor status, which were abstracted directly from patient medical records. Treatment data included surgery, radiation therapy and chemotherapy, all analyzed as dichotomous variables (received vs. not received). Also, information on type of surgery was collected. However, data was not available on treatment completion or dose reduction for chemotherapy and radiation therapy. Breast tumor hormone receptor status was considered positive if either ER or PR concentrations were >10 fmol per mg by the charcoal dextran assay. ER/PR status was categorized as: negative, positive and unknown. Cancer stage, based on pathological stage and in its absence on clinical stage, was analyzed according to the AJCC TNM staging system.[23] The cancer stage groups were treated as dichotomous indicator or "dummy" variables including stages II, III, IV and unstaged.

3.4.3 Anthropometric Data

Anthropometric measurements including weight and height were collected near or at time of breast cancer diagnosis by trained study staff at HFHS. Weight was determined using a scale. Subject's heights were measured without footwear, with a stadiometer. Weight and height were recorded in patient medical records. Later, data pertaining to weight and height were directly abstracted from patient medical records and converted to BMI.

3.4.4 Survival and Cause-of-Death Data

Survival follow-up and cause-of-death data were obtained from the HFHS and Metropolitan Detroit SEER tumor registries and from Michigan Department of Vital Statistics death certificate data. Classification of survival status into the categories alive, breast cancer death or competing cause death in the two registries had 91.9% agreement and where cause of death data were present in both registries agreement was high– the kappa statistic was 0.98.[90] The last date of follow-up was May 1, 2002.

3.4.5 Breast Cancer Recurrence/Progression Data

Data on failure to control cancer, as either local recurrence following resection or progression reflected in local expansion or regional or distant spread of cancer, were abstracted from patient medical records.

3.5 Analytic Strategy

Descriptive statistics were prepared using Fisher's exact test for categorical data, nonparametric test of trend for ordinal data, t-test for continuous data in two categories, ANOVA for continuous data with more than two categories and difference in follow-up using life table analysis (based on the log-rank test). The two goals of modeling were:

(1) Describe whether BMI was independently associated with breast cancer recurrence/progression and breast cancer-specific death, after covariate adjustment?

(2) Determine whether BMI modeled using restricted cubic splines improved prediction for breast cancer recurrence/progression and breast cancer-specific death?

Aim one reflected model development which focused on hypothesis testing in which adjustment for confounding and confidence intervals (or *p*-values) were important. In research, hypothesis testing frequently attempts to investigate causality or etiology and mechanisms of action. Understanding associations is useful for prevention and intervention purposes.[16]

Aim two was based on prediction model development. Prediction model development needs to be guided more by the models ability to discriminate and its calibration. Gail and Pfeiffer found that in screening applications discriminatory power was more important than calibration.[91] In modeling priority was given to improving discrimination, while at the same time attempting to optimize calibration and goodness-of-fit.

3.5.1 Modeling BMI Using Restricted Cubic Splines

It should not be assumed that all relationships involving one or more continuous variables are linearly associated. Some biological relationships are nonlinear and more complex. Restricted cubic splines are higher order polynomials used in exploring complex nonlinear relationships. Restricted cubic splines are superior to splines, cubic splines and quadratic terms because they are able to fit highly curving shapes.[15] As previously mentioned, quadratic terms are not versatile because they cannot describe sudden changes in shape and are unstable where data is thin. Also, quadratic terms are not capable of describing a wide range of nonlinear patterns.[89]

Position and number of knots: In choosing the position of the knots, placing knots at fixed percentiles is recommended. For example, if five knots are used, they can often be placed at the 5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles. In Stata, placing knots at

fixed percentiles is the default option unless specified. However, Stone found that knot location is unimportant. The fit depends on the number of knots utilized.[15] Choosing the number of knots depend on the sample size and variable distribution. Five knots or greater is rarely required in constructing restricted cubic splines. For majority of datasets, four knots are adequate and enable a valid comparison between flexibility and loss of precision caused by over fitting a small sample.[15]

In both Stata and R statistical programs, the two programs used in this study, Harrell's percentile approach was used. In Stata, the ***mkspline*** with options ***cubic*** and ***nknots*** command was utilized to construct restricted cubic splines for BMI, using four knots and three splines (k-1). In R statistical program, restricted cubic splines for BMI were created using the ***Design*** and ***Hmisc*** libraries produced by Harrell. Specifically, using the ***lrm*** command and specifying ***rcs (bmi, 4)***. The four knots were determined by the splines ability to sensibly describe the data based on data distribution percentiles.

3.5.2 *Research Question 1*: Are $BMI_{3splines}$ Independently Associated with Breast Cancer Outcomes?

A priori covariates chosen from past literature were: age, stage, ER/PR status, surgery, radiation therapy and chemotherapy. Potential confounders such as race, SES, menopausal status and family history were assessed in multivariate logistic regression. Confounding was assessed by comparing the crude and adjusted estimates of the measure of association. If the crude and adjusted measures differed 10-15%, then confounding was present. Also, interactions between $BMI_{3splines}$ * race and $BMI_{3splines}$ * menopausal status for breast cancer outcomes were assessed. Logistic regression was used to estimate odds ratios (OR), 95% confidence intervals (95% CI) and *p*-values for associations with

breast cancer recurrence/progression and breast cancer-specific death. However, because there was less misclassification of time of death and greater completeness of death data compared to recurrence/progression data, study conclusions were based on the death analysis. Survival analysis was carried out using Kaplan-Meier,[92] life table and Cox proportional hazards regression analyses.[93] Cox proportional hazards regression was used to estimate hazard ratios (HR), 95% CI and *p*-values for association with breast cancer-specific death. Since Cox regression uses time-to-event, breast cancer recurrence/progression was not assessed in Cox regression because date of recurrence/progression was not as accurate as the death data. Cox regression assumptions were evaluated graphically (using log minus log survival plots) and statistically (using the global test for trend) using Stata ***estat phtest, rank detail*** command which tests the change in scaled Schoenfeld residuals with time (*Appendix 1*).[94] Log minus log plots were created using Stata command ***stphplot***. Compared to logistic regression, Cox regression is more powerful in some ways because Cox regression uses time-to-event data which utilizes more information. However, Cox regression does not provide specific probabilities, which are useful for patients, clinicians and researchers. In this study, the HR and OR from modeling were used for comparative purposes and study conclusions were based on logistic regression (ORs).

The association between BMI three splines ($BMI_{3splines}$) and breast cancer recurrence/progression and breast cancer-specific death were evaluated using the likelihood ratio test (LRT), which compared the nested model with and without $BMI_{3splines}$. The LRT was performed using Stata ***lrtest*** command. $BMI_{3splines}$ were considered to be independently associated with breast cancer recurrence/progression and

breast cancer-specific death in logistic and Cox regression if LRT $p \leq 0.05$. All reported p- values were two-sided.

3.5.3 *Research Question 2:* Assessing the Nonlinear Relationship between BMI and Breast Cancer Outcomes

Locally weighted scatter plot smother or *lowess* is a graphical technique used to fit smooth curves to empirical data.[95] *Lowess* is nonparametric because the fitted curve is obtained empirically rather than through stringent prior specifications. *Lowess* plots reveal complex relationships that may be easily overlooked with traditional statistical modeling procedures such as linear regression. The *lowess* procedure is often conceptualized as a "vertical sliding window" that moves across the horizontal scale axis of the scatterplot.[95] The window stops and estimates a separate regression equation (using weighted least squares) at each of the evaluation points (or knots). Since the regressions only involve the data points that fall within the window, the estimated slopes can change to follow the contours of the data. In Stata, the default bandwidth is 0.80. Thus 80% of the data points that fall within each window are used to estimate the regression equations. This feature gives *lowess* the flexibility to conform to relatively complicated, nonlinear shapes within the point cloud of a scatterplot. In Stata, the ***lowess*** command was used to construct the figures.

Lowess plots were constructed to graphically describe the relationship between BMI and breast cancer outcomes. Also, *lowess* plots were created to describe the relationship between BMI and breast cancer outcomes with advanced and unknown stage excluded. This was done because it is believed that those women who lose or lost weight

(low BMI) did so because of their advanced disease. *Lowess* figures with and without advanced and unknown stage were compared for similarities and differences.

To determine the significance of the nonlinear component between BMI and breast cancer outcomes R statistical program was used with the specific library ***Hmisc***. The ***anova*** command yielded the Wald test p-value for nonlinearity.

3.5.4 *Research Question 3*: Model Performance (Discrimination, Calibration and Validation)

Breast cancer models were built using a "backward" elimination technique. Breast cancer covariates, chosen *a prior*, predicting breast cancer recurrence/progression and breast cancer-specific death included: age, stage, ER/PR status, surgery, radiation therapy, chemotherapy, race, SES, menopausal status, family history and $BMI_{3splines}$. Initially, all covariates were included in multivariate logistic regression. Covariates with the highest p-value were dropped from breast cancer models. This process was repeated until all covariates with $p \geq 0.2$ were eliminated. Multivariate logistic regression was used to determine if $BMI_{3splines}$ improved the ability to predict breast cancer outcomes. The receiver operator characteristic area under the curve (ROC AUC) in logistic regression and Harrell's concordance or c-statistic in Cox regression were used to assess the discriminatory ability of the breast cancer models. A ROC curve plots the sensitivity against the false-positive rate (i.e. one minus the specificity) for a range of thresholds. ROC curves for tests with overall good performance will lie close to the left and topmost margins of the plot. Stata ***comproc*** command was used to test whether there was significant improvement in discrimination for $BMI_{3splines}$. ***Comproc*** tests the null hypothesis that the ROC AUC in the model including $BMI_{3splines}$ is equal to the ROC

AUC in the model excluding $BMI_{3splines}$, where models are nested, i.e. include exactly the same individuals. ***Comproc*** yields bootstrap standard errors and bias-corrected (BC) 95% CIs. Bootstrap resampling for ***comproc*** was performed using 1000 replications. The c-statistic takes into account the proportion of all usable subject pairs in which the predictions and outcomes are concordant. Then, the c-statistic ranks the hazard, i.e. of death, and determines whether this hazard corresponds with those patients who lived longer with the outcome. In assessing the c-statistic, a c-statistic of 0.5 is equivalent to random or chance discrimination whereas a c-statistic of 1.0 indicates perfect discrimination. According to Harrell, a c-statistic > 0.8 is useful in predicting the responses of individual subjects. Furthermore, these two measures of discriminatory ability have been shown to be mathematically equivalent.[96, 97] Harrell's c-statistic was performed using Stata ***estat concordance*** command. Furthermore, in R statistical program using ***validate*** and ***calibrate***, Harrell's uncorrected and bootstrap optimism-corrected c-statistic was calculated using Somer's optimism-corrected D_{xy} [c-statistic= $(D_{xy}/2) + 0.5$]. Bootstrapping, an internal validation technique, estimates the sampling distribution of an estimator by resampling with replacement from the original sample. The difference in the apparent and test validation yields the bootstrap optimism-corrected value. Bootstrap optimism-corrected values were performed using 200 replications.[15] Somer's D_{xy} looks at the difference between concordant and discordant probabilities. A D_{xy} of 0 indicates that the model is making random predictions whereas a D_{xy} of 1 indicates perfectly discriminating predictions. It has been demonstrated that Somer's D_{xy} and Harrell's c-statistic are linearly related and are a measure of a models discriminatory ability.[16]

In addition, overall model performance of the logistic regression models predicting breast cancer outcomes were assessed using Nagelkerke's pseudo R^2 and the Hosmer-Lemeshow goodness-of-fit test. The Hosmer-Lemeshow goodness-of-fit test divides subjects into deciles based on predicted probabilities and then computes a chi-square from the observed and expected frequencies. A Hosmer-Lemeshow goodness-of-fit test $p \geq 0.05$ implies that the models estimates fit the data at an acceptable level. In Stata, the ***estat gof, group (10)*** command was used to calculate the Hosmer-Lemeshow goodness-of-fit-test.

Finally, calibration assesses how well the predicted probabilities match the observed probabilities. Calibration was assessed using the calibration slope, mean absolute error and 90th percentile absolute error. The slope summarizes the deviation of model predicted probabilities from observed proabbilities.[16] The error in mean absolute error and 90th percentile absolute error refers to the difference between the ideal and the corresponding bias-corrected calibrated values. In assessing calibration, the three major components are: the slope, mean absolute error and 90th percentile absolute error. Thus, a larger slope and a smaller error value indicate a well calibrated model. Also, calibration plots were created plotting predicted probabilities vs. observed probabilities. Three lines define a calibration plot: (1) The ideal line is a 45° line representing a perfectly calibrated model with a slope of one, (2) The apparent line is the bootstrap uncorrected line depicting the breast cancer models, and (3) The bias-corrected line is the bootstrap optimism-corrected line.[15] Calibration was evaluated in R statistical program using the ***Design*** and ***Hmisc*** libraries produced by Harrell with the specific package ***calibrate***.

Stata/IP 10.1 (Stata Corporation, College Station, Tex) and R (version 2.11.1) statistical programs were used to prepare models, statistics and figures.

CHAPTER 4: RESULTS

4.1 Population Demographics

An initial 906 subjects were included in the study. However, outcome data were available for 897 subjects. The population characteristics are provided in *Table 4* describing the HFHS study population.

***Table 4.* Characteristics of the 897 Individuals from HFHS (Detroit, Michigan)**

Characteristics		HFHS patients (N= 897)
Age in years (SD; range)		60.7 (14.5; 24-94)
Race	White	634 (70.7%)
	Black	263 (29.3%)
Menopausal status	Premenopausal	178 (19.8%)
	Postmenopausal	719 (80.2%)
Family history	No	388 (43.3%)
	Yes	235 (26.2%)
	Unknown	274 (30.5%)
Follow-up	Median (range) in years	10.0 (0.04-17.8)
Recurrence/progression	No	518 (57.7%)
	Yes	281 (31.3%)
	Unknown	98 (10.9%)
Deaths	No	418 (46.6%)
	Breast cancer	180 (20.1%)
	Competing causes	299 (33.3%)
BMI (kg/m^2) (SD; range) (N= 816)		27.2 (6.3; 15.1-55.8)
Stage	Carcinoma in situ	30 (3.3%)
	I	282 (31.4%)
	II	378 (42.1%)
	III	91 (10.1%)
	IV	48 (5.4%)
	Unstaged	68 (7.7%)
Estrogen/progesterone receptor	Negative	114 (12.7%)
	Positive	514 (57.3%)
	Unknown	269 (30.0%)
Surgery	No	40 (4.5%)
	Yes	818 (91.2%)
	Unknown	39 (4.3%)
Radiation	No	563 (62.8%)
	Yes	256 (28.5%)
	Unknown	78 (8.7%)
Chemotherapy	No	566 (63.1%)
	Yes	245 (27.3%)
	Unknown	86 (9.6%)
Tamoxifen	No	469 (52.3%)
	Yes	346 (38.6%)
	Unknown	82 (9.1%)

Abbreviations: BMI, body mass index; N, sample size; SD, standard deviation.

The mean age at breast cancer diagnosis was 60.7 years and mean BMI, at diagnosis, was 27.2 kg/m^2. Majority (76.8%) of patients were diagnosed with early stage breast cancer (Stage 0-II). ER/PR status was as follows: 114 (12.7%) patients were ER/PR negative whereas 514 (57.3%) patients were ER/PR positive. Eight hundred eighteen (91.2%) patients underwent surgery whereas 40 (4.5%) did not. Also, 256 (28.5%) received radiation therapy compared to 563 (62.8%) of patients who did not. Finally, 245 (27.3%) received chemotherapy whereas 566 (63.1%) did not. The median follow-up was 10 years (range 0.04-17.8). During follow-up, 281 patients experienced a breast cancer recurrence/progression and 180 breast cancer-specific deaths were reported.

The distribution of breast cancer recurrence/progression and breast cancer status by baseline characteristics for selected variables is presented in *Tables 5* and *6*, respectively.

Table 5. Breast Cancer Recurrence/Progression in HFHS Patients (Detroit, Michigan), by Baseline Characteristics (N= 799)

Characteristics	No (N= 518)	Yes (N= 281)	*P* Value*
Age in years (SD; range)	60.4 (14.4; 24-93)	60.5 (14.8; 26-94)	0.89
Race			
White	376 (67.6%)	180 (32.4%)	0.02
Black	142 (58.4%)	101 (41.6%)	
Menopausal status			
Premenopausal	109 (66.5%)	55 (33.5%)	0.65
Postmenopausal	409 (64.4%)	226 (35.6%)	
Family history			
No	243 (68.1%)	114 (31.9%)	0.93
Yes	147 (67.4%)	71 (32.6%)	
Follow-up			
Median (range) in years	10.8 (0.04-17.8)	5.1 (0.04-17.6)	< 0.001
BMI (kg/m^2) (SD; range) (N= 742)	27.3 (6.2; 16.3-53.8) (N= 487)	27.1 (6.7; 15.1-55.8) (N= 255)	0.64
Stage			
Carcinoma in situ	24 (100%)	0	
I	212 (84.5%)	39 (15.5%)	
II	227 (65.2%)	121 (34.8%)	< 0.001
III	29 (34.1%)	56 (65.9%)	
IV	2 (4.3%)	44 (95.7%)	
Unstaged†	24 (53.3%)	21 (46.7%)	0.11
Estrogen/progesterone receptor			
Negative	54 (49.5%)	55 (50.5%)	< 0.001
Positive	306 (67.4%)	148 (32.6%)	
Surgery			
No	9 (23.7%)	29 (76.3%)	< 0.001
Yes	508 (67.6%)	243 (32.4%)	
Radiation			
No	368 (70.4%)	155 (29.6%)	< 0.001
Yes	135 (57.0%)	102 (43.0%)	
Chemotherapy			
No	382 (73.6%)	137 (26.4%)	< 0.001
Yes	117 (50.0%)	117 (50.0%)	
Tamoxifen			
No	312 (71.6%)	124 (28.4%)	< 0.001
Yes	191 (58.8%)	134 (41.2%)	

Abbreviations: BMI, body mass index; N, sample size; SD, standard deviation.

* P values were computed for dichotomous variables using the Fisher exact test, for age and BMI using the *t* test, for follow-up using life-table analysis (based on the log-rank test) and stage using the nonparametric test for trend.

† Unstaged is not included in the nonparametric test for trend; assessing whether a significant difference exists in unstaged breast cancer between those women who did not and did experience a breast cancer recurrence/progression.

As depicted in *Table 5,* there was no significant difference in age between those subjects who did not experience breast cancer recurrence/progression from those who did. Subjects who did not experience breast cancer recurrence/progression had a significantly longer follow-up (median 10.8 years) when compared to subjects who experienced recurrence/progression from breast cancer (median 5.1 years; log-rank test $p < 0.001$ testing the difference of the entire survival experience of both groups). There was no significant difference in BMI between subjects who did and did not experience breast cancer recurrence/progression. A significant trend was observed for breast cancer recurrence/progression across the ordered levels of stage ($p < 0.001$). There was a significant difference in ER/PR status between subjects without and with breast cancer recurrence/progression ($p < 0.001$). Finally, there was a significant difference in surgery, radiation therapy, chemotherapy and tamoxifen between subjects without and with recurrence/progression from breast cancer ($p < 0.001$).

Table 6. Death Status in HFHS Patients (Detroit, Michigan), by Baseline Characteristics (N= 897)

Characteristics	Alive (N= 418)	Competing causes death (N= 299)	Breast cancer-specific death (N= 180)	*P* Value*
Age in years (SD; range)	55.0 (12.5; 24-83)	69.1 (13.3; 26-94)	60.1 (14.0; 30-90)	< 0.001
Race				
White	316 (49.8%)	203 (32.0%)	115 (18.2%)	0.006
Black	102 (38.8%)	96 (36.5%)	65 (24.7)	
Menopausal status				
Premenopausal	123 (69.1%)	11 (6.2%)	44 (24.7%)	< 0.001
Postmenopausal	295 (41.0%)	288 (40.1%)	136 (18.9%)	
Family history				
No	201 (51.8%)	114 (29.4%)	73 (18.8%)	0.39
Yes	126 (53.6%)	58 (24.7%)	51 (21.7%)	
Follow-up				
Median (range) in years	12.7 (0.08-17.8)	6.3 (0.04-16.4)	3.3 (0.05-15.7)	< 0.001
BMI (kg/m^2) (SD; range) (N= 816)	27.1 (5.8; 16.3-53.8) (N= 389)	27.6 (6.7; 16.3-53.3) (N= 269)	26.6 (6.9; 15.1-55.8) (N= 158)	0.31
Stage				
Carcinoma in situ	22 (73.3%)	8 (26.7%)	0	
I	167 (59.2%)	94 (33.3%)	21 (7.5%)	
II	177 (46.8%)	124 (32.8%)	77 (20.4%)	< 0.001
III	20 (21.9%)	29 (31.9%)	42 (46.2%)	
IV	1 (2.1%)	21 (43.7%)	26 (54.2%)	
Unstaged†	31 (45.6%)	23 (33.8%)	14 (20.6%)	0.97
Estrogen/progesterone receptor				
Negative	45 (39.5%)	19 (16.7%)	50 (43.8%)	< 0.001
Positive	239 (46.5%)	164 (31.9%)	111 (21.6%)	
Surgery				
No	3 (7.5%)	23 (57.5%)	14 (35.0%)	< 0.001
Yes	399 (48.8%)	262 (32.0%)	157 (19.2%)	
Radiation				
No	266 (47.3%)	205 (36.4%)	92 (16.3%)	< 0.001
Yes	121 (47.3%)	65 (25.4%)	70 (27.3%)	
Chemotherapy				
No	262 (46.3%)	228 (40.3%)	76 (13.4%)	< 0.001
Yes	125 (51.0%)	39 (15.9%)	81 (33.1%)	
Tamoxifen				
No	244 (52.0%)	141 (30.1%)	84 (17.9%)	0.01
Yes	143 (41.3%)	125 (36.1%)	78 (22.6%)	

Abbreviations: BMI, body mass index; N, sample size; SD, standard deviation.

* *P* values were computed for categorical variables using contingency table analysis (and Fisher exact test), for age and BMI using ANOVA, for follow-up using life-table analysis (based on the log-rank test) and stage using the nonparametric test for trend.

† Unstaged is not included in the nonparametric test for trend; assessing whether a significant difference exists in unstaged breast cancer between those women who survived, died due to competing causes and due to breast cancer-specific death.

In evaluating death status by baseline characteristics, there was a significant difference in age between those who survived, died due to competing causes and died due to breast cancer-specific death. Those who died from breast cancer had a shorter follow-up (median 3.3 years) compared to those who died from competing causes (median 6.3 years; log-rank test $p < 0.001$ testing the difference of the entire survival experience of both groups). There was no significant difference in BMI between those subjects who survived, died of competing causes and died due to breast cancer-specific death. A significant trend was observed between those who survived, died of competing causes and died due to breast cancer-specific death across the ordered levels of stage ($p < 0.001$). In assessing hormone receptor status, there was a significant difference in ER/PR status between those subjects who survived, died due to competing causes and died from breast cancer-specific death ($p < 0.001$). Finally, there was a significant difference in surgery, radiation therapy and chemotherapy between those subjects who survived, died of competing causes and died due to breast cancer-specific death ($p < 0.001$).

The distribution of baseline characteristics for selected variables by BMI categories are presented in *Table 7.*

Table 7. Baseline Characteristics of the HFHS (Detroit, Michigan) Study Population, by BMI Categories (N= 816)

Characteristics‡	< 22 kg/m² (N= 163)	22-25 kg/m² (N= 194)	> 25-30 kg/m² (N= 246)	≥ 30 kg/m² (N= 213)	P Value*
Age in years (SD; range)	59.2 (16.3; 24-94)	60.1 (15.5; 31-92)	61.6 (12.8; 30-93)	61.4 (13.2; 28-88)	0.32
Race					
White	134 (82.2%)	157 (80.9%)	166 (67.5%)	121 (56.8%)	< 0.001
Black	29 (17.8%)	37 (19.1%)	80 (32.5%)	92 (43.2%)	
Menopausal status					
Premenopausal	40 (24.5%)	50 (25.8%)	37 (15.0%)	34 (16.0%)	0.01
Postmenopausal	123 (75.5%)	144 (74.2%)	209 (85.0%)	179 (84.0%)	
Family history					
No	72 (55.0%)	99 (65.6%)	110 (62.2%)	101 (64.3%)	0.28
Yes	59 (45.0%)	52 (34.4%)	67 (37.8%)	56 (35.7%)	
Follow-up					
Median (range) in years	8.6 (0.05-17.0)	11.2 (0.04-17.8)	10.8 (0.08-17.4)	9.8 (0.25-17.7)	0.27
Stage					
Carcinoma in situ	4 (2.4%)	12 (6.2%)	7 (2.8%)	4 (1.8%)	
I	55 (33.7%)	72 (37.1%)	82 (33.3%)	59 (27.7%)	
II	65 (39.9%)	73 (37.7%)	109 (44.3%)	105 (49.3%)	0.16
III	15 (9.2%)	21 (10.8%)	23 (9.3%)	24 (11.3%)	
IV	13 (8.0%)	7 (3.6%)	12 (5.0%)	13 (6.1%)	
Unstaged†	11 (6.8%)	9 (4.6%)	13 (5.3%)	8 (3.8%)	0.61
Estrogen/progesterone receptor					
Negative	26 (23.4%)	23 (16.7%)	24 (14.2%)	28 (19.4%)	0.24
Positive	85 (76.6%)	115 (83.3%)	145 (85.8%)	116 (80.6%)	
Surgery					
No	10 (6.2%)	5 (2.6%)	7 (2.9%)	13 (6.2%)	0.13
Yes	152 (93.8%)	186 (97.4%)	235 (97.1%)	198 (93.8%)	
Radiation					
No	111 (71.6%)	120 (64.9%)	151 (66.5%)	144 (70.9%)	0.43
Yes	44 (28.4%)	65 (35.1%)	76 (33.5%)	59 (29.1%)	
Chemotherapy					
No	104 (67.1%)	131 (73.2%)	161 (70.9%)	134 (67.0%)	0.50
Yes	51 (32.9%)	48 (26.8%)	66 (29.1%)	66 (33.0%)	
Tamoxifen					
No	87 (56.5%)	109 (59.9%)	136 (60.2%)	107 (52.2%)	0.33
Yes	67 (43.5%)	73 (40.1%)	90 (39.8%)	98 (47.8%)	
Death status					
Alive	66 (40.5%)	104 (53.6%)	122 (49.6%)	97 (45.5%)	
Competing causes	54 (33.1%)	54 (27.8%)	84 (34.1%)	77 (36.2%)	0.08
Breast cancer	43 (26.4%)	36 (18.6%)	40 (16.3%)	39 (18.3%)	

Abbreviations: BMI, body mass index; N, sample size; SD, standard deviation.

* *P* values were computed for categorical variables using contingency table analysis (Fisher exact test), for age and BMI using ANOVA, for follow-up using life-table analysis (based on the log-rank test) and stage using the nonparametric test for trend.

† Unstaged is not included in the nonparametric test for trend; assessing whether a significant difference exists in unstaged breast cancer between BMI categories.

‡ In this table column percentages are reported.

As depicted in *Table 7,* there was no significant difference in age, family history, follow-up, stage, ER/PR status, surgery, radiation therapy, chemotherapy and tamoxifen and death status between those women who were underweight, normal weight, overweight and obese near or at breast cancer diagnosis. However, there was a significant difference in race and menopausal status between those women who were underweight, normal weight, overweight and obese near or at breast cancer diagnosis ($p < 0.001$ and $p= 0.01$, respectively)

4.2 Covariates Included in the Breast Cancer Recurrence/Progression Model

4.2.1 Unadjusted Logistic Regression

A priori covariates used in the breast cancer recurrence/progression model are presented in *Table 8.*

Table 8. **Unadjusted Logistic Regression Odds Ratios for Breast Cancer Recurrence/Progression (N=799)**

Characteristics		HFHS patients
Age (per 10 years)		1.01 (0.91-1.11; p= 0.89)
Stage	Carcinoma in situ + stage I	Baseline
	Stage II	3.23 (2.15-4.83; $p < 0.001$)
	Stage III	11.69 (6.66-20.50; $p < 0.001$)
	Stage IV	133.13 (31.01-571.52; $p < 0.001$)
	Unstaged	5.29 (2.69-10.42; $p < 0.001$)
Estrogen/progesterone receptor	Negative	Baseline
	Positive	0.47 (0.31-0.73; $p < 0.001$)
	Unknown	0.48 (0.31-0.77; p= 0.002)
Treatment	Surgery (N= 789)	0.15 (0.07-0.32; $p < 0.001$)
	Radiation (N= 760)	1.79 (1.30-2.47; $p < 0.001$)
	Chemotherapy (N= 753)	2.79 (2.02-3.85; $p < 0.001$)
BMI (N= 742)	Spline 1*	0.81 (0.71-0.93; p= 0.002)
	Spline 2*	2.38 (1.25-4.52; p= 0.008)
	Spline 3*	0.13 (0.03-0.66; p= 0.013)

Abbreviations: BMI, body mass index; N, sample size.
* Knot location for $BMI_{3splines}$: 19.22 kg/m^2, 23.96 kg/m^2, 28.29 kg/m^2 and 38.44 kg/m^2.

Stage, ER/PR status, surgery, radiation therapy and chemotherapy were associated with breast cancer recurrence/progression in unadjusted analysis ($p < 0.001$). However, age was not associated with breast cancer recurrence/progression in unadjusted logistic regression (p= 0.89).

4.2.2 Multivariate Logistic Regression

In *Table 9,* multivariate logistic regression independent variables effect estimates, 95% CIs and p-values are presented for breast cancer recurrence/progression.

Table 9. **Multivariate Logistic Regression Odds Ratios and Performance Statistics for Breast Cancer Recurrence/Progression (N= 692)**

Characteristics		HFHS patients
Age (per 10 years)		1.24 (1.06-1.44; p= 0.01)
Stage	Carcinoma in situ + stage I	Baseline
	Stage II	3.43 (2.10-5.60; $p < 0.001$)
	Stage III	10.50 (5.41-20.39; $p < 0.001$)
	Stage IV	95.36 (21.11-430.80; $p < 0.001$)
	Unstaged	2.75 (1.09-6.93; p= 0.03)
Estrogen/progesterone receptor	Negative	Baseline
	Positive	0.62 (0.36-1.08; p= 0.09)
	Unknown	0.58 (0.32-1.04; p= 0.07)
Treatment	Surgery	0.20 (0.06-0.60; p= 0.004)
	Radiation	2.04 (1.36-3.08; p= 0.001)
	Chemotherapy	2.43 (1.51-3.94; $p < 0.001$)
BMI	Spline 1*	0.77 (0.65-0.92; p= 0.003)
	Spline 2*	3.00 (1.35-6.68; p= 0.01)
	Spline 3*	0.07 (0.01-0.52; p= 0.01)
MODEL PERFORMANCE		
LRT p-value for $BMI_{3splines}$		p= 0.0230
Nonlinear p-value		p= 0.0194
ROC AUC w & w/o $BMI_{3splines}$		0.8006; 0.7929
Comparison of ROC summary statistics w & w/o $BMI_{3splines}$		AUC delta= 0, p= 0.14
R^2		0.364†; 0.322‡
Hosmer-Lemeshow goodness-of-fit-test		p= 0.699
c-statistic		0.801; 0.782‡
Calibration		
	Calibration slope	0.883‡
	Mean absolute error	0.018
	90th percentile absolute error	0.038

Abbreviations: BMI, body mass index; LRT, likelihood ratio test; N, sample size; ROC AUC, receiver operator characteristic area under the curve.
* Knot location for $BMI_{3splines}$: 19.22 kg/m^2, 23.96 kg/m^2, 28.29 kg/m^2 and 38.44 kg/m^2.
† Nagelkerke's R^2.
‡ Bootstrap optimism-corrected R^2, c-statistic and calibration slope obtained using Harrell's validate and calibrate function in R.
Note: Race, SES, menopausal status and family history were tested in the breast cancer recurrence/progression model and deemed unimportant and are therefore excluded from modeling.

In multivariate logistic regression, age was significantly associated with breast cancer recurrence/progression (p= 0.01). However, age was not significant in unadjusted analysis. Stage was highly associated with breast cancer recurrence/progression in both unadjusted and multivariate analysis ($p < 0.001$). ER/PR status was not significant in multivariate logistic regression however results were trending to significance. On the other hand, ER/PR status was significant in unadjusted logistic regression. Finally, surgery, radiation therapy and chemotherapy remained significant in both unadjusted and multivariate analysis.

4.2.2.1 Testing of Relevant Covariates for Breast Cancer Recurrence/Progression

Covariates: Aside from the well-established breast cancer covariates, other variables from past literature were tested including race, menopausal status and SES. When race was added to the multivariate model in *Table 9,* it was not significant (p= 0.08). However, it was trending to significance. Similarly, menopausal status and SES in multivariate logistic regression were not significantly associated with breast cancer recurrence/progression (p= 0.96 and p= 0.89, respectively).

Interaction terms: Possible interaction terms between $BMI_{3splines}$ * race and $BMI_{3splines}$ * menopausal status were evaluated. $BMI_{3splines}$ * race and $BMI_{3splines}$ * menopausal status as interaction terms were not significantly associated with breast cancer recurrence/progression (p= 0.50 and p= 0.12, respectively).

4.2.3 *Research Question 1:* $BMI_{3splines}$ and Breast Cancer Recurrence/Progression

According to the LRT (*Table 9*), $BMI_{3splines}$ were independently associated with breast cancer recurrence/progression (p= 0.023).

4.2.4 *Research Question 2:* Assessing the Nonlinear Association between $BMI_{3splines}$ and Breast Cancer Recurrence/Progression

A *lowess* curve describing the nonlinear relationship between $BMI_{3splines}$ and breast cancer recurrence/progression is presented in *Figure 12.* Also, a *lowess* curve describing the nonlinear relationship between $BMI_{3splines}$ and breast cancer recurrence/progression with advanced and unknown stage excluded is presented in *Figure 13.*

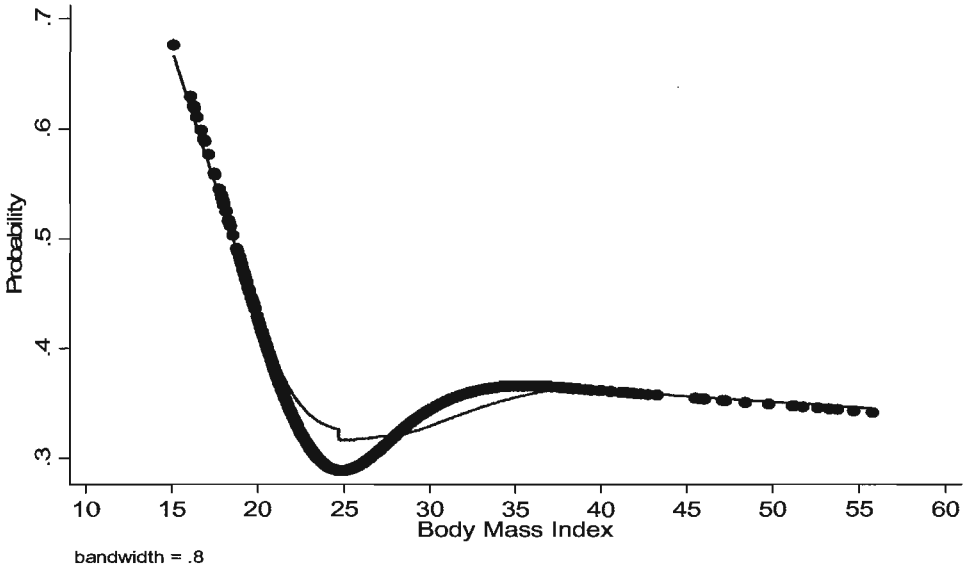


***Figure 12. Lowess* Curve Describing the Relationship between $BMI_{3splines}$ and Breast Cancer Recurrence/Progression in the HFHS (Detroit, Michigan) Data**

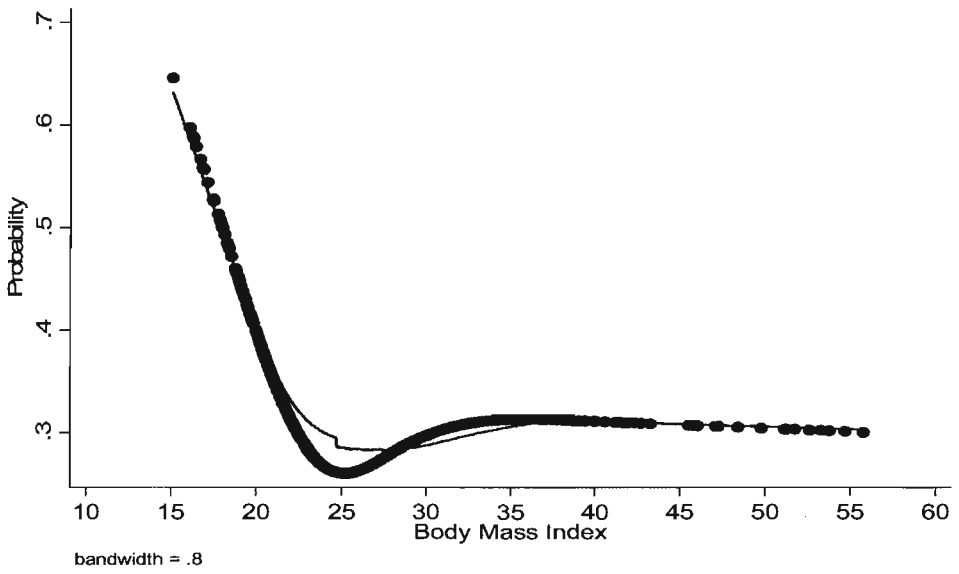


***Figure 13. Lowess* Curve Describing the Relationship between $BMI_{3splines}$ and Breast Cancer Recurrence/Progression with Advanced and Unknown Stage Excluded**

According to *Figure 12,* the probability of breast cancer recurrence/progression increased with decline in BMI from roughly 25 kg/m^2 and increased with increase in BMI from the above mentioned BMI. The increase in probability was evident from a BMI of 25-35 kg/m^2. Then a leveling off and a possible slight decline in the probability of breast cancer recurrence/progression with increasing BMI was observed. The increase in probability appeared to be higher for low BMI as opposed to high BMI. This effect might have been explained away by those patients with advanced stage breast cancer losing weight (BMI) due to their advanced disease. For this reason analysis was limited to early stage breast cancer cases (< stage III) only (*Figure 13)*; the nonlinear effect appeared to be remarkably similar in shape to *Figure 12.* The Wald test *p*-value assessing nonlinearity between BMI and breast cancer recurrence/progression was $p= 0.02$; indicating a significant nonlinear association between BMI and breast cancer recurrence/progression.

4.2.5 *Research Question 3:* Testing Model Performance for the Model Predicting Breast Cancer Recurrence/Progression

As depicted in *Table* 9, the bootstrap bias-corrected (BC) estimates assessing the ROC AUC in the nested model with and without $BMI_{3splines}$ were 0.8006 (bootstrap SE 0.02; BC 95% CI= 0.76-0.83) and 0.7929 (bootstrap SE 0.02; BC 95% CI= 0.76-0.83). Although not statistically significant at the 0.05 level, the results are trending to significance (p= 0.14) and are suggestive of a possible difference in the ROC AUCs. A ROC AUC comparing the nested model with and without $BMI_{3splines}$ is presented in *Figure 14.*

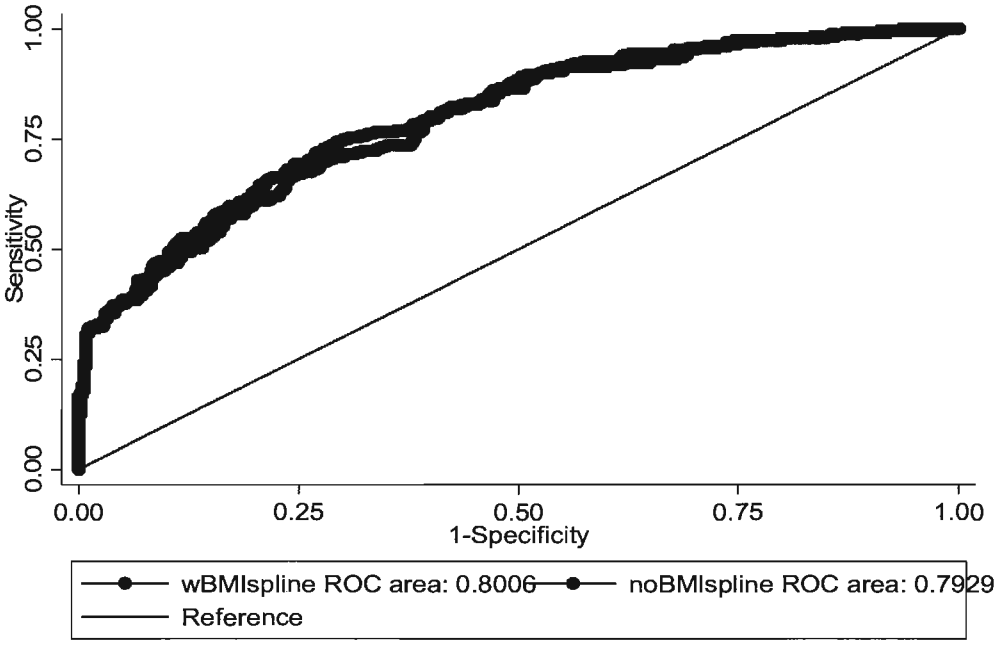


Figure 14. **ROC Curve for Breast Cancer Recurrence/Progression in the Nested Model with and without $BMI_{3splines}$ in the HFHS (Detroit, Michigan) Data**

The uncorrected and bootstrap optimism-corrected c-statistics predicting breast cancer recurrence/progression were 0.801 and 0.782. As a result, the model predicting breast cancer recurrence/progression demonstrated good discriminatory ability between

those subjects who did not and did experience a recurrence/progression from breast cancer. The model predicting breast cancer recurrence/progression had a pseudo R^2 of 0.364 and a bootstrap optimism-corrected pseudo R^2 of 0.322. Furthermore, the Hosmer-Lemeshow goodness-of-fit test *p*-value was 0.699, implying that the model's estimates fit the data at an acceptable level.

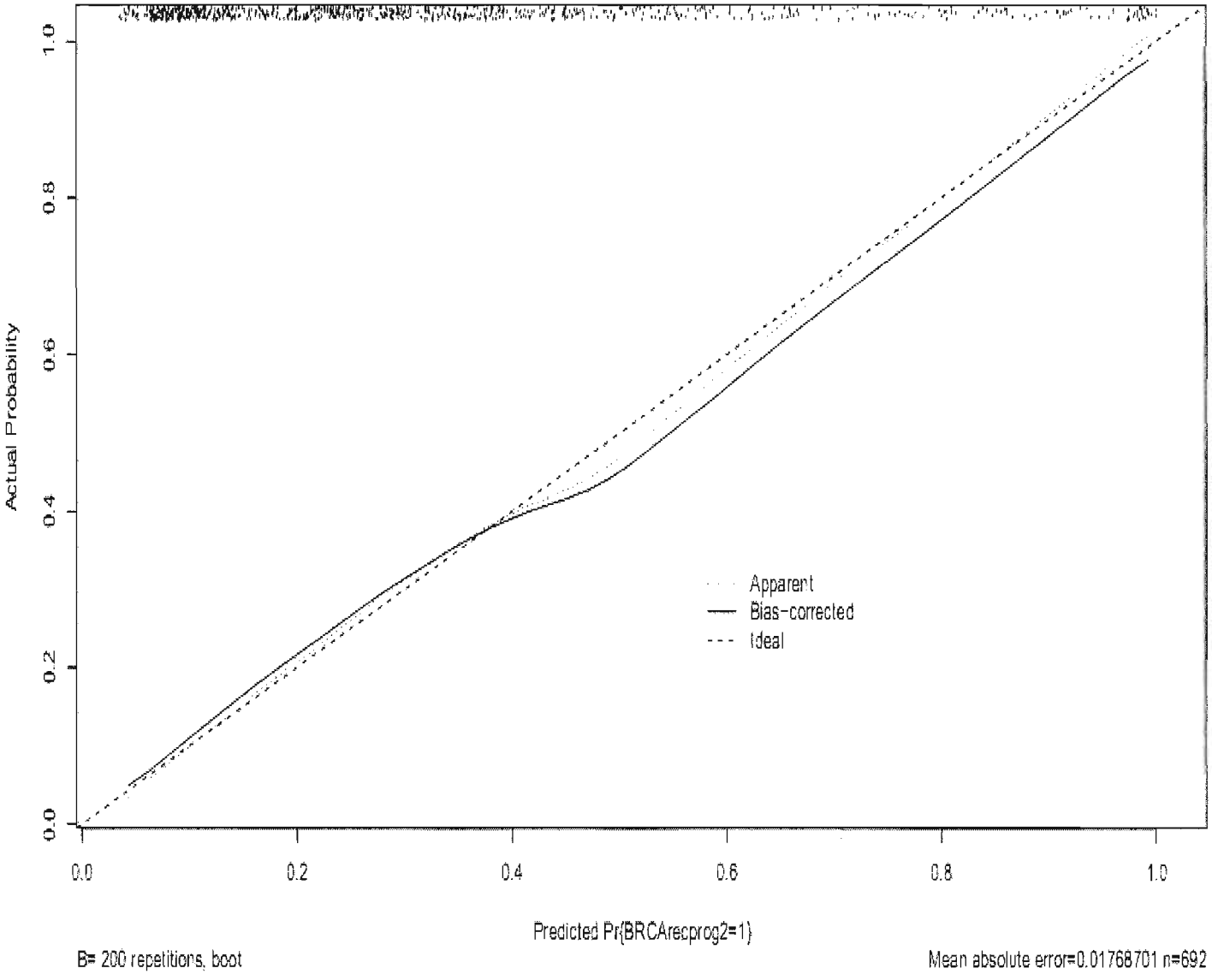


Figure 15. **Calibration Plot Predicting Breast Cancer Recurrence/Progression**

A calibration plot (plotting the predicted probabilities vs. observed probabilities) is presented in *Figure 15.* In assessing the calibration plot, the calibration slope of the prognostic model is 0.883. Also, the mean absolute error (0.018) and 90th percentile absolute error (0.038) are small. This indicates that the model predicting breast cancer recurrence/progression is well calibrated as measured by the calibration slope, mean absolute error and 90th percentile absolute error.

4.3 Inclusion of Covariates in the Breast Cancer-Specific Death Model

4.3.1 Unadjusted Logistic Regression

A priori covariates from previous literature in unadjusted logistic regression are presented in *Table 10.*

Table 10. **Unadjusted Logistic Regression Odds Ratios for Breast Cancer-Specific Death (N= 897)**

Characteristics		HFHS patients
Age (per 10 years)		0.97 (0.86-1.08; p= 0.54)
Stage	Carcinoma in situ + stage I	Baseline
	Stage II	3.54 (2.13-5.90; $p < 0.001$)
	Stage III	11.88 (6.49-21.75; $p < 0.001$)
	Stage IV	16.38 (7.97-33.65; $p < 0.001$)
	Unstaged	3.59 (1.72-7.50; p= 0.001)
Estrogen/progesterone receptor	Negative	Baseline
	Positive	0.35 (0.23-0.54; $p < 0.001$)
	Unknown	0.10 (0.05-0.18; $p < 0.001$)
Treatment	Surgery (N= 858)	0.44 (0.23-0.86; p= 0.02)
	Radiation (N= 819)	1.93 (1.35-2.75; $p < 0.001$)
	Chemotherapy (N= 811)	3.18 (2.22-4.56; $p < 0.001$)
BMI (N= 816)	Spline 1*	0.84 (0.73-0.96; p= 0.01)
	Spline 2*	1.77 (0.89-3.52; p= 0.10)
	Spline 3*	0.30 (0.05-1.67; p= 0.17)

Abbreviations: BMI, body mass index; N, sample size.
* Knot location for $BMI_{3splines}$: 19.22 kg/m^2, 23.96 kg/m^2, 28.29 kg/m^2 and 38.44 kg/m^2.

In unadjusted analysis, age was not significantly associated with breast cancer-specific death (p= 0.54). Stage was highly associated with breast cancer-specific death. Patients in stage IV of their disease had 16.38-fold greater odds of dying from breast cancer compared to patients with carcinoma in situ and stage I ($p < 0.001$). In assessing hormone receptor status, ER/PR status was significantly associated with breast cancer-specific death. Subjects with ER/PR positive hormone status had 0.35-fold lower odds of dying from breast cancer compared to patients with ER/PR negative hormone status ($p <$ 0.001). Surgery, radiation therapy and chemotherapy were significantly associated with breast cancer-specific death.

4.3.2 Multivariate Logistic Regression

Well established covariates associated with breast cancer-specific death were forced into multivariate logistic regression. Effect estimates, associated 95% CIs and p-values are presented in *Table 11*.

Table 11. **Multivariate Logistic Regression Odds Ratios and Performance Statistics for Breast Cancer-Specific Death (N= 743)**

Characteristics		HFHS patients
Age (per 10 years)		1.32 (1.10-1.58; p= 0.002)
Stage	Carcinoma in situ + stage I	Baseline
	Stage II	3.40 (1.79-6.46; $p < 0.001$)
	Stage III	11.31 (5.16-24.78; $p < 0.001$)
	Stage IV	18.82 (7.23-48.98; $p < 0.001$)
	Unstaged	1.92 (0.48-7.76; p= 0.36)
Estrogen/progesterone receptor	Negative	Baseline
	Positive	0.53 (0.30-0.93; p= 0.03)
	Unknown	0.09 (0.04-0.18; $p < 0.001$)
Treatment	Surgery	1.12 (0.38-3.30; p= 0.84)
	Radiation	2.31 (1.44-3.69; $p < 0.001$)
	Chemotherapy	3.86 (2.24-6.66; $p < 0.001$)
BMI	Spline 1*	0.78 (0.66-0.93; p= 0.01)
	Spline 2*	2.21 (0.95-5.14; p= 0.07)
	Spline 3*	0.18 (0.02-1.51; p= 0.12)
MODEL PERFORMANCE		
LRT p-value for $BMI_{3splines}$		p= 0.0101
Nonlinear p-value		p= 0.0121
ROC AUC w & w/o $BMI_{3splines}$		0.8424; 0.8331
Comparison of ROC summary statistics w & w/o $BMI_{3splines}$		AUC delta= 0, p= 0.08
R^2		0.374†; 0.327‡
Hosmer-Lemeshow goodness-of-fit-test		p= 0.210
c-statistic		0.842; 0.824‡
Calibration		
Calibration slope		0.892‡
Mean absolute error		0.025
90th percentile absolute error		0.050

Abbreviations: BMI, body mass index; LRT, likelihood ratio test; N, sample size; ROC AUC, receiver operator characteristic area under the curve.

* Knot location for $BMI_{3splines}$: 19.22 kg/m^2, 23.96 kg/m^2, 28.29 kg/m^2 and 38.44 kg/m^2.

† Nagelkerke's R^2.

‡ Bootstrap optimism-corrected R^2, c-statistic and calibration slope obtained using Harrell's validate and calibrate function in R.

Note: Race, SES, menopausal status and family history were tested in the breast cancer-specific death model and deemed unimportant and are therefore excluded from modeling.

Age, in the multivariate model (*Table 11),* was significantly associated with breast cancer-specific death. For every 10 year increase in age those women diagnosed with breast cancer had a 32% greater odds of dying from breast cancer (p= 0.002). Stage adjusted for all covariates was significantly associated with breast cancer-specific death. Women with stage IV disease had 18.82-fold greater odds of dying from breast cancer compared to women who had carcinoma in situ and stage I ($p < 0.001$). ER/PR positive hormone status compared to ER/PR negative hormone status demonstrated a protective effect of breast cancer-specific death (OR= 0.53; p= 0.03). Surgery was not significantly associated with breast cancer-specific death (p= 0.84) whereas radiation therapy and chemotherapy were significantly adversely associated with breast cancer-specific death ($p < 0.001$).

4.3.2.1 Testing of Additional Covariates for Breast Cancer-Specific Death

Covariates: Additional sociodemographic and clinical covariates deemed important were assessed in multivariate logistic regression. Inclusions of covariates were guided by past literature. Race, adjusted for age, stage, ER/PR status, surgery, radiation therapy, chemotherapy and $BMI_{3splines}$ was not significantly associated with breast cancer-specific death (p= 0.24). Similarly, SES, adjusted for the above mentioned covariates was not significantly associated with breast cancer-specific death (p= 0.98). Finally, menopausal status was not significantly associated with breast cancer-specific death (p= 0.67).

Interaction terms: Interaction terms between $BMI_{3splines}$ * race and $BMI_{3splines}$ * menopausal status were created and tested for breast cancer-specific death. $BMI_{3splines}$ * race and $BMI_{3splines}$ * menopausal status were not significantly associated with breast cancer-specific death (p= 0.48 and p= 0.25, respectively).

4.3.3 Covariates Included in Multivariate Cox Regression for Breast Cancer-Specific Death

In assessing breast cancer-specific death, a multivariate Cox regression model is presented in *Table 12*.

Table 12. **Multivariate Cox Regression Hazard Ratios for Breast Cancer-Specific Death (N= 743)**

Characteristics		HFHS patients
Age (per 10 years)		1.31 (1.13-1.52; $p < 0.001$)
Stage	Carcinoma in situ + stage I	Baseline
	Stage II	3.37 (1.88-6.04; $p < 0.001$)
	Stage III	7.81 (4.10-14.88; $p < 0.001$)
	Stage IV	29.38 (14.05-61.45; $p < 0.001$)
	Unstaged	2.02 (0.58-7.07; $p= 0.27$)
Estrogen/progesterone receptor	Negative	Baseline
	Positive	0.65 (0.43-0.98; $p= 0.04$)
	Unknown	0.26 (0.15-0.47; $p < 0.001$)
Treatment	Surgery	1.46 (0.66-3.25; $p= 0.35$)
	Radiation	2.27 (1.58-3.28; $p < 0.001$)
	Chemotherapy	2.61 (1.72-3.95; $p < 0.001$)
BMI	Spline 1*	0.75 (0.66-0.86; $p < 0.001$)
	Spline 2*	2.87 (1.48-5.857; $p= 0.002$)
	Spline 3*	0.09 (0.02-0.48; $p= 0.01$)
MODEL PERFORMANCE		
LRT p-value for $BMI_{3splines}$		$p= 0.0003$
c-statistic		0.82

Abbreviations: BMI, body mass index; N, sample size.
* Knot location for $BMI_{3splines}$: 19.22 kg/m^2, 23.96 kg/m^2, 28.29 kg/m^2 and 38.44 kg/m^2.
Note: Race, SES, menopausal status and family history were tested in the breast cancer-specific death model and deemed unimportant and are therefore excluded from modeling.

Age, adjusted for the aforementioned covariates was significantly associated with breast cancer-specific death ($p < 0.001$). Stage was highly associated with breast cancer-specific death. Women with stage IV disease had 29.38-fold greater hazard of death from breast cancer compared to women who had carcinoma in situ and stage I ($p < 0.001$). ER/PR positive status was associated with breast cancer-specific death ($p= 0.04$). Finally, surgery was not significantly associated with breast cancer-specific death whereas

radiation therapy and chemotherapy were highly adversely associated with breast cancer-specific death ($p < 0.001$).

The Cox regression model was used to determine if effect estimates were consistent with logistic regression effect estimates. Effect estimates and standard errors were consistent between Cox and logistic regression models. For this reason, study conclusions are based on logistic models as presented.

4.3.4 *Research Question 1:* $BMI_{3splines}$ and Breast Cancer-Specific Death in Multivariate Logistic Regression

According to the LRT (*Table 11)*, $BMI_{3splines}$ were independently associated with breast cancer-specific death ($p= 0.0101$).

4.3.5 *Research Question 2:* Assessing the Nonlinear Association between $BMI_{3splines}$ and Breast Cancer-Specific Death

A *lowess* curve describing the nonlinear relationship between $BMI_{3splines}$ and breast cancer-specific death is presented in *Figure 16.* A *lowess* curve describing the nonlinear relationship between $BMI_{3splines}$ and breast cancer-specific death with advanced and unknown stage excluded is presented in *Figure 17.*

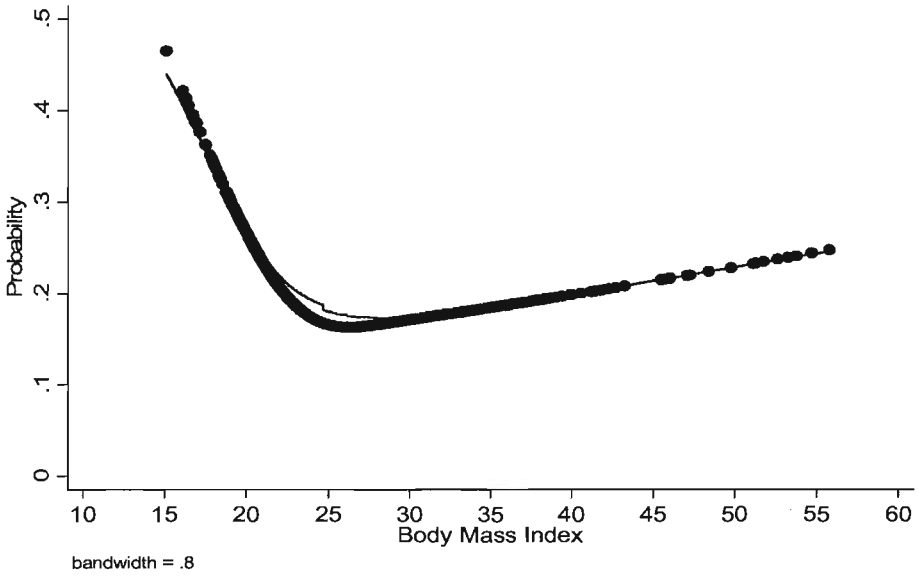


***Figure 16. Lowess* Curve Describing the Relationship between $BMI_{3splines}$ and Breast Cancer-Specific Death in the HFHS (Detroit, Michigan) Data**

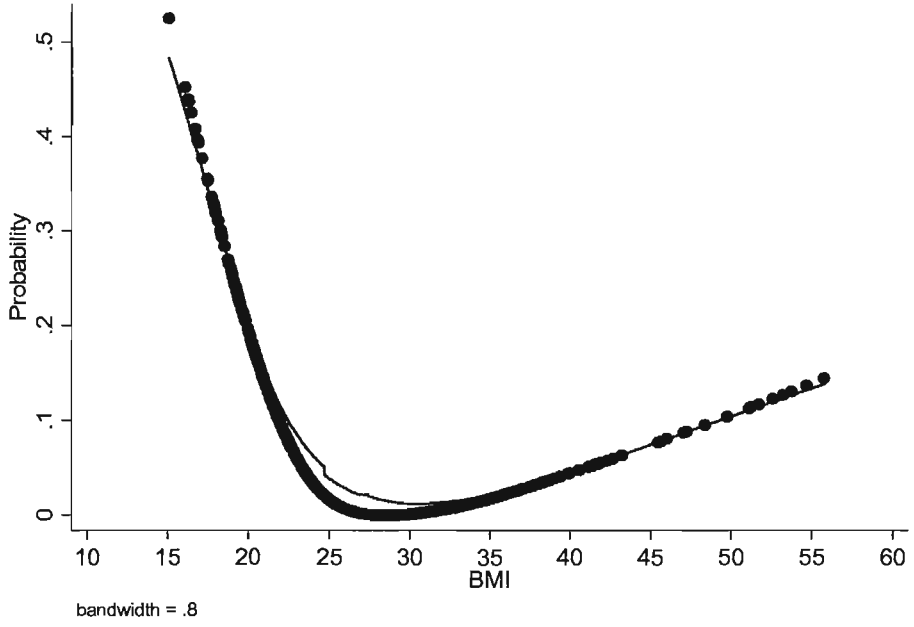


***Figure 17. Lowess* Curve Describing the Relationship between $BMI_{3splines}$ and Breast Cancer-Specific Death with Advanced and Unknown Stage Excluded**

As depicted in *Figure 16,* the probability of breast cancer-specific death increased with decline in BMI from roughly 25 kg/m^2 and increased with increase in BMI from 25 kg/m^2. The increase in probability appeared to be higher for low BMI as opposed to high BMI. This effect might have been explained away by those patients with advanced stage breast cancer losing weight (BMI) due to their advanced disease. For this reason analysis was limited to early stage (< stage III) breast cancer cases only (*Figure 17).* The nonlinear effect appeared to be remarkably similar in shape to *Figure 16.* The Wald test *p*-value assessing nonlinearity between BMI and breast cancer-specific death was $p=$ 0.01; indicating a significant nonlinear association between BMI and breast cancer-specific death.

4.3.6 *Research Question 3:* Testing Model Performance for the Model Predicting Breast Cancer-Specific Death

As presented in *Table 11*, the bootstrap bias-corrected ROC AUC estimates assessing the difference in the nested model with and without $BMI_{3splines}$ were 0.8424 (bootstrap SE 0.02; BC 95% CI= 0.81-0.87) and 0.8331 (bootstrap SE 0.02; BC 95% CI= 0.80-0.87). Although not statistically significant at the 0.05 level, the results are trending to significance ($p=$ 0.08) and are suggestive of a possible difference in the ROC AUCs. A ROC AUC comparing the nested model with and without $BMI_{3splines}$ for breast cancer-specific death is presented in *Figure 18.*

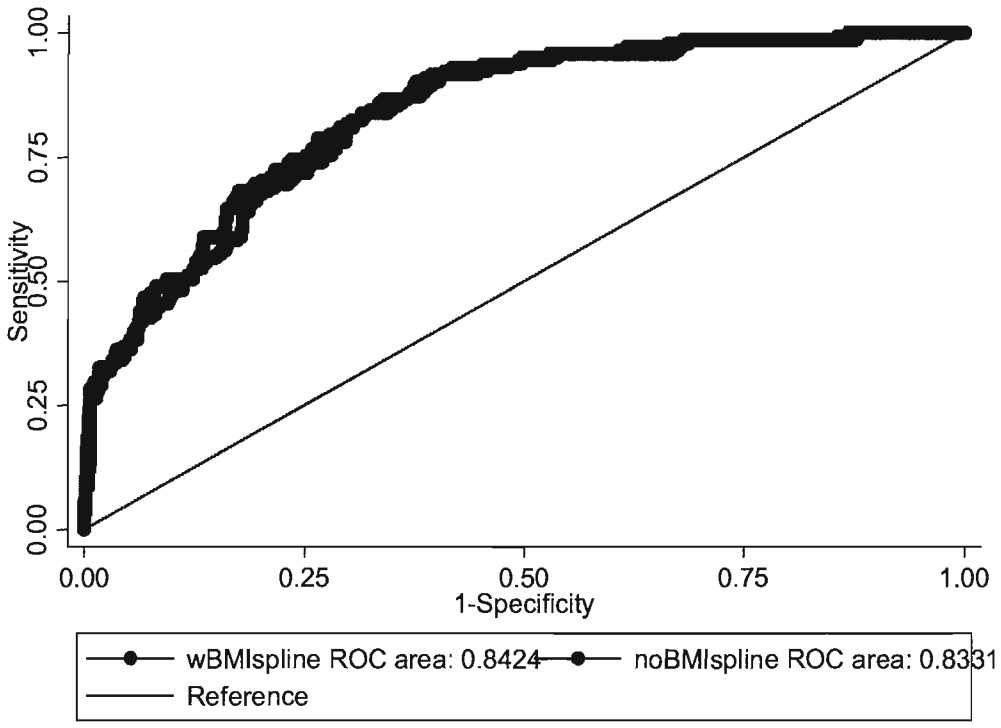


Figure 18. **ROC Curve for Breast Cancer-Specific Death in the Nested Model with and without $BMI_{3splines}$ in the HFHS (Detroit, Michigan) Data**

The uncorrected and bootstrap optimism-corrected c-statistics predicting breast cancer-specific death were 0.842 and 0.824. The model predicting breast cancer-specific death demonstrated good discriminatory ability between those subjects who did not and did experience a death from breast cancer. The model predicting breast cancer-specific death had a Nagelkerke's pseudo R^2 of 0.374 and a bootstrap optimism-corrected Nagelkerke's pseudo R^2 of 0.327. Furthermore, the Hosmer-Lemeshow goodness-of-fit test *p*-value was 0.210, implying that the model's estimates fit the data at an acceptable level.

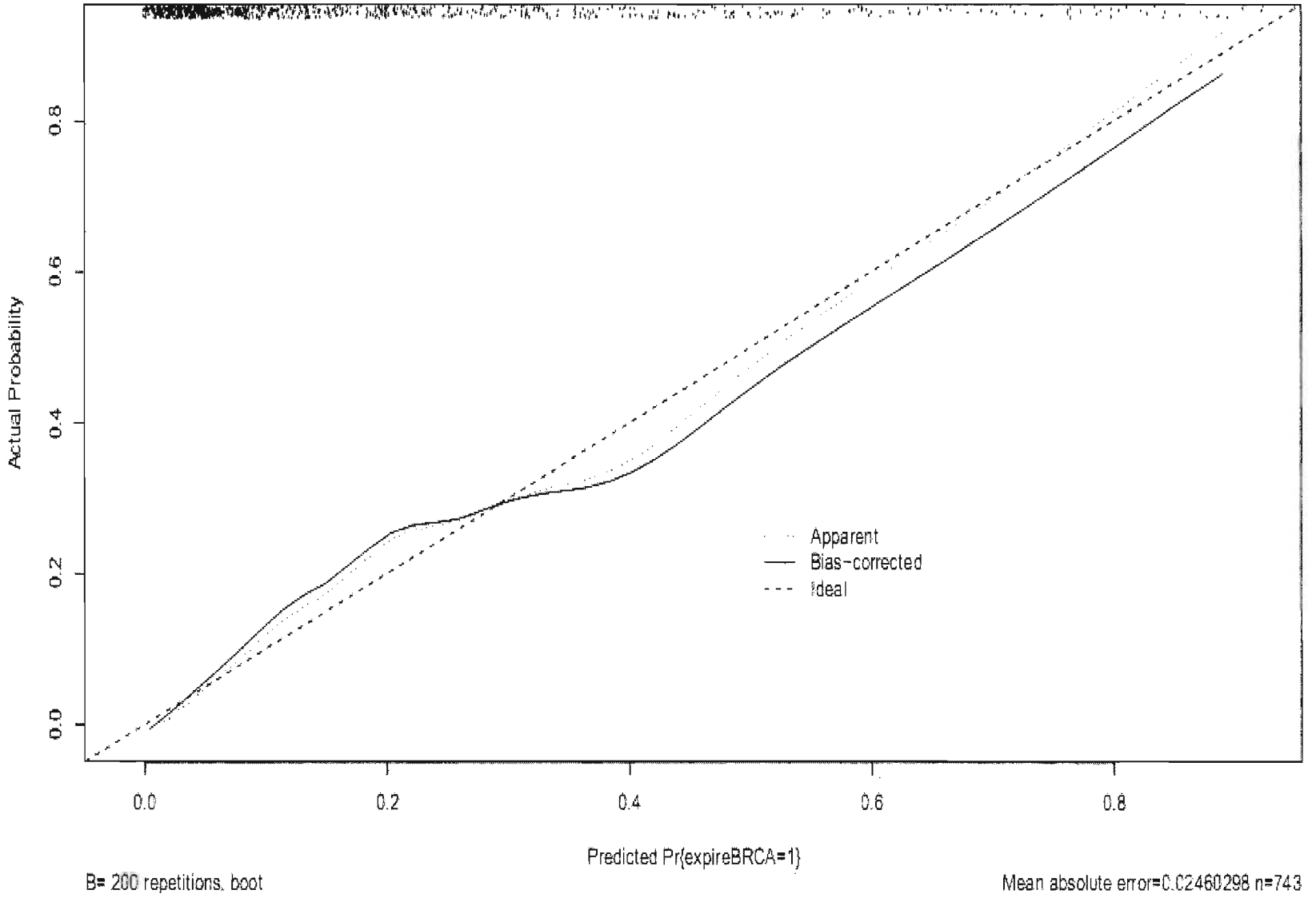


Figure 19. **Calibration Plot Predicting Breast Cancer-Specific Death**

A calibration plot (plotting the predicted probabilities vs. observed probabilities) is presented in *Figure 19*. In assessing the calibration plot, the calibration slope of the breast cancer-specific death prognostic model is 0.892. Also, the mean absolute error (0.025) and 90th percentile absolute error (0.050) are relatively small. This indicates that the model predicting breast cancer-specific death is reasonably well calibrated as measured by the calibration slope, mean absolute error and 90th percentile absolute error.

4.3.7 $BMI_{3splines}$ and Breast Cancer-Specific Death in Multivariate Cox Regression

As depicted in *Table* 12, $BMI_{3splines}$ were highly associated with breast cancer-specific death (LRT p= 0.0003). The c-statistic for the model predicting breast cancer-specific death was c= 0.82.

CHAPTER 5: DISCUSSION & CONCLUSIONS

5.1 Pertinent Study Findings

5.1.1 The Association between $BMI_{3splines}$ and Breast Cancer Outcomes

Adjusted for age, stage, ER/PR status, surgery, radiation therapy and chemotherapy, $BMI_{3splines}$ near or at breast cancer diagnosis were independently associated with breast cancer recurrence/progression and breast cancer-specific death (LRT p= 0.023 and LRT p= 0.010, respectively).

5.1.2 The Nonlinear Relationship between $BMI_{3splines}$ and Beast Cancer Outcomes

The probability of breast cancer recurrence/progression and breast cancer-specific death increased with decline in BMI from roughly 25 kg/m^2 and increased with increase in BMI from 25 kg/m^2. A J-shaped relationship best described the relationship between BMI and breast cancer recurrence/progression and breast cancer-specific death. The Wald test p-value assessing nonlinearity between BMI and breast cancer recurrence/progression and breast cancer-specific death were p= 0.02 and p= 0.01, respectively. This indicated a significant nonlinear relationship between BMI and breast cancer recurrence/progression and breast cancer-specific death.

5.1.3 Breast Cancer Prediction Models

The ROC AUCs for breast cancer recurrence/progression and breast cancer-specific death were not statistically significant. However, results were trending to significance and were suggestive of a possible difference in the ROC AUCs, in the nested model with and without $BMI_{3splines}$ (p= 0.14 and p= 0.08). The c-statistic demonstrated good discriminatory ability between those subjects who did not and did experience a breast

cancer recurrence/progression (uncorrected c-statistic= 0.801 and bootstrap optimism-corrected c-statistic= 0.782) and those who did not and did die from breast cancer (uncorrected c-statistic= 0.842 and bootstrap optimism-corrected c-statistic= 0.824). The goodness-of-fit test implied that breast cancer recurrence/progression and breast cancer-specific death estimates fit the data at an acceptable level (p= 0.210 and p= 0.699, respectively). In assessing calibration, the bootstrap optimism-corrected calibration slopes are relatively large for breast cancer recurrence/progression and breast cancer-specific death (0.883 and 0.892) whereas the mean absolute errors (0.018 and 0.025) and 90th percentile absolute errors (0.038 and 0.050) are relatively small for breast cancer recurrence/progression and breast cancer-specific death, respectively. As a result, models predicting breast cancer recurrence/progression and breast cancer-specific death are well calibrated.

5.2 Past Study Findings

5.2.1 Independent Relationship between Nonlinearly Modeled BMI and Breast Cancer Outcomes

Our results demonstrated a nonlinear association between $BMI_{3splines}$ near or at breast cancer diagnosis and breast cancer recurrence/progression and breast cancer-specific death. Suissa and colleagues found a similar nonlinear association however BMI was modeled using quadratic terms. Also, BMI was nonlinearly associated with breast cancer recurrence and overall survival.[87] Then Suissa and colleagues' results were validated by Goodwin and colleagues.[71] Similarly, Goodwin and colleagues demonstrated that BMI, modeled using quadratic terms, represented a significant improvement over the linear

model. Likewise, a significant association between BMI and distant recurrence and overall survival persisted.[71]

As previously mentioned, quadratic terms are limited in that they can only describe a limited number of nonlinear patterns and estimates are unstable in extreme ranges such as the tails.[89] However, this study modeled BMI using restricted cubic splines as this method is deemed superior[89]. That is because restricted cubic splines are able to describe a greater range of nonlinear patterns and thus fit the data better.[16, 88]

5.2.2 Describing the Nonlinear Relationship between BMI and Breast Cancer Outcomes

Suissa and colleagues demonstrated that the probability of overall survival was concave and lowest at a BMI of 34 kg/m^2.[87] A BMI of 34-50 kg/m^2 was associated with the highest probability of overall survival. However, the probability of overall survival was highest for high vs. low BMI.[87] Furthermore, Goodwin and colleagues found a similar J-shaped relationship and demonstrated that women with either low or high BMI (< 20 kg/m^2 or >25 kg/m^2) had the highest hazard of distant recurrence and overall survival.[71] Similar to Suissa and colleagues' work, the effect was highest for high vs. low BMI.

Contrary to the risk picture demonstrated by Suissa and colleagues and Goodwin and colleagues[71, 87], in this study, models with restricted cubic splines conveyed that the probability was highest for low vs. high BMI. Results from this study indicate and support a J-shaped relationship between BMI and breast cancer recurrence/progression and breast cancer-specific death. It was postulated that this effect might have been explained away by those patients with advanced stage breast cancer losing weight (low BMI) due to their advanced disease. However, this explanation was not adequate because

when analysis was limited to early stage breast cancer cases only (< stage III); the nonlinear effect appeared to be similar in shape.

5.2.3 Low BMI and Breast Cancer Outcomes

Studies also found an association between underweightness (low BMI) and breast cancer prognosis. Moon and colleagues found that adjusted for age, tumor size, lymph node metastasis, ER/PR status, grade and lymphovascular invasion underweight patients had a significantly lower overall survival and breast cancer-specific survival.[61] Also, underweight women had a significantly higher risk of local recurrence and distant metastasis.[61] In addition, Chen and colleagues concluded that low BMI and young age were poor prognostic factors for locoregional recurrence in multivariate analysis.[98] Similarly Marret and colleagues' work found that low BMI was an independent predictive factor of local recurrence after breast conserving therapy.[99] To date, majority of studies have focused on the relationship between high BMI and breast cancer prognosis and ignored the impact of low BMI and breast cancer prognosis. Failure to recognize such a subgroup may affect interpretations of previous study results. In addition, this study helped more accurately characterize the probability/risk of adverse outcomes in breast cancer patients- to a level not previously achieved.

5.3 Study Limitations

5.3.1 Study Power

Study power was possibly an issue because it may have been low. If the sample size were larger than breast cancer models would have better prediction. To prevent overfit in prediction modeling, 10-20 outcomes per predictor are required in model development.

5.3.2 Misclassification Bias

There may be more misclassification of time of recurrence/progression and less completeness of recurrence/progression data compared to death data.

5.3.3 Missing Data

The proportion of patients used in the breast cancer recurrence/progression and breast cancer-specific death models was 76.4% and 82.0%. Some of the covariates used in modeling such as surgery, radiation therapy, chemotherapy and $BMI_{3splines}$ contained missing data. Patterns of missing values are reported in *Appendix 2.* Future analyses may consider using multiple imputation methods as simulation studies have demonstrated that imputing data can be superior compared to using complete case analysis.[100]

5.3.4 BMI as a Proxy Measure for Obesity

BMI is a proxy measure for obesity however BMI does not discriminate between body fat and lean body mass. A study conducted by Loi and colleagues demonstrated a high correlation between BMI and body fat ($r= 0.91; p < 0.001$).[80] These findings were reported in an Asian population. Results might be different depending on the ethnic variation but for the most part in clinical practice and research, BMI is convenient and simple while demonstrating a high correlation with percent body fat.[79]

5.3.5 Absence of Genetic Mutation Data

Models predicting breast cancer recurrence/progression and breast cancer-specific death did not include genetic mutation data. Beenken and colleagues reported that adjusted for age, stage, nodal status, grade and ER/PR status, p53 tumor suppressor gene mutation and c-erB-2 overexpression have independent prognostic significance with breast cancer recurrence and overall survival.[101] Overexpression of c-erB-2 and p53 tumor suppressor

gene mutation correlated strongly with poor patient survival. After a 16 year follow-up, Beenken and colleagues suggested that coexpression of c-erB-2 and p53 tumor suppressor gene mutation may have more prognostic significance than traditional prognostic factors such as stage and nodal status.[101] Similarly, Allred and colleagues suggested a prognostic role for p53 tumor suppressor gene mutation.[102] Likewise, Lai and colleagues found that adjusted for stage, treatment and number of mutations, patients with p53 tumor suppressor gene mutations had significantly greater breast cancer-specific deaths compared to patients without p53 tumor suppressor gene mutations.[103] Furthermore, the presence of a dose-response relationship between the number of mutations and the likelihood of dying from breast cancer provided evidence of a causal relationship between p53 tumor suppressor gene mutations and breast cancer disease-free survival.[103] Mechanistically, it has been postulated that breast cancers co-expressing c-erB-2 and p53 tumor suppressor gene mutations have lost a key mechanism for control of cell proliferation and have gained an activator of malignant cell potential, resulting in a highly malignant tumor phenotype.[101] According to the literature future prognostic models need to combine the predictive power of individual molecular biomarkers with specific clinical and pathologic factors.

The HFHS comorbidity dataset did not contain acquired mutation data however germline mutation data such as family history was available and is considered a surrogate measure for an inherited genetic component. Since minimal genetic data was available this study did not incorporate a genetic aspect. Family history was assessed in unadjusted and multivariate logistic regression for breast cancer recurrence/progression and breast cancer-specific death. In unadjusted logistic regression family history was not

significantly associated with breast cancer recurrence/progression (OR= 1.03; 95% CI= 0.72-1.48; p= 0.87). In multivariate analysis adjusted for age, stage, ER/PR status, surgery, radiation therapy, chemotherapy and $BMI_{3splines}$, family history remained insignificant (OR= 1.19; 95% CI= 0.75-1.88; p= 0.46). Similarly, family history was not significantly associated with breast cancer-specific death in unadjusted and multivariate analysis, adjusted for age, stage, ER/PR status, surgery, radiation therapy, chemotherapy and $BMI_{3splines}$, (OR= 1.20; 95% CI= 0.80-1.79; p= 0.38 and OR= 1.38; 95% CI= 0.83-2.30; p= 0.22, respectively). As a result, adding family history to the association study would be irrelevant because family history was not significantly associated with breast cancer outcomes. Furthermore, by definition, a confounder must be associated with the exposure under study. To date, we are unaware of evidence indicating associations between p53 and c-erB-2 and BMI. On the other hand, incorporating tumour-specific genetic data to the prediction study might be relevant and results may or may not be affected because to improve prediction very strong effect estimates (high ORs) are needed.

5.4 Study Strengths

5.4.1 Study Design

The strength of this study was the study design utilized. The historical cohort design is valid because exposure data is collected before the outcome occurs therefore there is no recall bias. Also, because it is a cohort, with naturally occurring exposures and outcomes, it is not vulnerable to selection bias the way a case-control study would have been. Furthermore, the historical cohort design is convenient because a preexisting cohort of incident breast cancer cases is used to collect exposure data (BMI) and the outcomes

(breast cancer recurrence/progression and breast cancer-specific death) are ascertained at the time the study is initiated. For this reason, results are collected sooner because the study has begun with a preexisting cohort to reduce the duration of the study.

5.4.2 Analytic Approaches

Using restricted cubic splines as an approach to studying nonlinear effects between BMI and breast cancer recurrence/progression and breast cancer-specific death strengthened the results of this study.

5.4.3 Discrimination, Calibration and Validation

Sophisticated statistical techniques such as discrimination and calibration were utilized to assess breast cancer prognostic models. Furthermore, results were internally validated using bootstrapping. Empirical and simulation studies indicate that bootstrapping internal validation is more efficient and potentially encompasses less bias than using the cross-validation approaches, where the sample is split into model development and testing subsets.[16, 88] To date, research has not implemented sophisticated statistical modeling techniques such as discrimination and calibration to adequately assess the predictive ability of breast cancer recurrence/progression and breast cancer-specific death models.

5.4.4 Completeness of Cause-of-Death Data

Survival follow-up and cause-of-death data were collected from SEER tumor registries and death certificates which are generally thought to be an accurate source of data collection since the exact date of death is specified. In comparing death certificates to hospital and autopsy records, death certificates are deemed as a more valid source for mortality data. As well, the US is required by law to have a death registry therefore the data are known to be complete.[104]

5.4.5 Inclusion of Early and Late Stage Breast Cancer Patients

Previous studies only took early stage breast cancer cases into account. This study included early and late stage breast cancer cases.

5.5 Biological Plausibility

5.5.1 Biological Explanations for Low BMI

Possible biological rationales that may explain the risk picture observed in this study and that past studies reported have been suggested for low and high BMI. The mechanisms between BMI and breast cancer prognosis are complex and are probably different for low and high BMI. However, there have been several biological rationales explaining the impact of low BMI on local recurrence but these hypotheses are not well established.[98, 99]

Timing of weight gain is considered a breast cancer risk factor. Women who progressively gain weight from puberty to menopause are considered to be at risk for developing breast cancer.[105] Another possible explanation is high endogenous estrogen levels observed in premenopausal lean women. Premenopausal lean women have higher estradiol levels compared to overweight women, while BMI is positively related to estradiol in postmenopausal women.[98]

Moon and colleagues suggest additional mechanisms that explain the association between low BMI and breast cancer prognosis.[61] One such explanation is the interaction between tumor cells and circulating immune cells through various molecular signals, from initial carcinogenesis to metastasis.[61] Immune cells may inhibit or promote tumor progression and influence the efficacy of systemic antitumor treatments. In addition, patients showing severe undernutrition and micronutrient deficiency, cytokine reactions and subsequent activation of the immune system are compromised, affecting tumor-

immune system interactions. As a result, this may be associated with an increased risk of breast cancer recurrence.[61] Finally, another possible mechanism linking underweightness and increased risk of breast cancer recurrence are the tumor-modulating roles of local systemic adipocytes.[61] Animal experiments demonstrate that the presence of mammary adipocytes is critical for mammary gland development and irradiation of mammary fat pads causes malignant progression of normal mammary cells; suggesting a protective role for mammary adipocytes.[61]

5.5.2 Biological Explanations for High BMI

The biological rationale explaining high BMI is clearer and more established. The biological effect of obesity on prognosis is associated with excessive endogenous estrogen production. This effect is strongest in those women who are postmenopausal. In premenopausal women, estrogen levels influenced by peripheral aromatization in adipose tissue would be presumably not significant compared with that produced by functioning ovaries. Even after covariate adjustment, obesity remained to be a significant predictor of breast cancer recurrence and all cause mortality on both ER/PR tumors; suggesting other biological influences aside from estrogen may be responsible for this effect.[80]

In obese premenopausal women, the hormonal make-up is different and obesity is associated with hyperandrogenism, hyperinsulinemia and lower serum-binding hormone globulin.[80] For example, hyperinsulinemia can be a potential mediator of the adverse effect of obesity in premenopausal breast cancer patients.[80] As a result, there is biologic rationale for an adverse prognostic effect of insulin.[71] Approximately, 90% of breast cancer cells express IGF-1, IGF-2, insulin and hybrid insulin/IGF-1 receptors. IGF-1 and insulin receptors, mediate the mitogenic effects of insulin, are frequently overexpressed

in breast cancer patients, with levels of insulin receptor that are 6-10-fold higher than in normal breast epithelium.[71]

5.6 Clinical and Public Health Implications

In research, hypothesis testing or causal modeling is important because if significant results are reported it enables the researcher to further investigate the etiology and mechanisms of action that might be driving the association. Therefore, it allows the researcher to determine the underlying cause that is affecting the exposure/disease relationship. If direct causal mechanisms are identified, randomized clinical trials are needed to see if dietary programs and exercise interventions or both may reduce elevated BMI and increase low BMI in order to improve breast cancer outcomes.

Prognostic models are crucial in clinical settings because physicians and health policy makers are required to make predictions on patient survival. Physicians use prediction models in their decision-making on screening and treatment of disease in high risk groups, diagnostic work-up (i.e. ordering a risky or expensive test) and choice of therapy. Clinical prediction models have the potential to provide the evidence-based input for shared decision-making, by providing estimates of the individual probabilities/risks.[16] Additionally, prognostic models are utilized by public health officials to assess disease burden. The health care system utilize prediction models to determine how much a patient is costing the health care system, as well as, help to develop plans to address patient needs.[16]

5.7 Future Research

First, BMI, which attempts to capture adiposity, is weighted to the square of height and this leads to low height individuals having adiposity overestimated in many individuals.

As a result, future studies may want to consider dissecting the findings by strata of height.

Second, study results are in need of external validation using independent study samples in order to validate the study results presented in this study. In addition, once findings are externally validated, superior models need to be built incorporating BMI but also additional markers excluded from our study because of absence of data, i.e. genetic tumor markers and functional status need to be included in future breast cancer prognostic models.

Finally, to implement breast cancer prediction models in clinical practice future studies need to construct nomograms for breast cancer recurrence/progression and breast cancer-specific death. These nomograms can be utilized by physicians and patients to generate predictions for patient survival. A nomogram predicting breast cancer recurrence/progression and breast cancer-specific death and directions on filling out the nomogram is presented in *Appendix 3*.

5.8 Conclusions

The results of this study indicate that adjusted for all prognostic covariates, $BMI_{3splines}$ near or at breast cancer diagnosis were independently associated with breast cancer-specific death. According to the Wald test for nonlinearity, BMI was nonlinearly associated with breast cancer-specific death. A J-shaped relationship best described the association between BMI and breast cancer-specific death. In addition, the probability of death due to breast cancer was highest for low vs. high BMI.

Discrimination as measured by ROC AUC, although not statistically significant but trending to significance was suggestive of a possible difference in the ROC AUCs,

between the nested model with and without $BMI_{3splines}$. Furthermore, the Hosmer-Lemeshow goodness-of-fit test demonstrated that the breast cancer-specific death estimates fit the data at an acceptable level. Finally, the model including $BMI_{3splines}$ to predict breast cancer-specific death was well calibrated as measured by the calibration slope, mean absolute error and 90^{th} percentile absolute error.

Overall, this new knowledge may aid in the revision of pre-existing nutritional and exercise programs or may help in the development of new interventions that are specifically geared towards women diagnosed with breast cancer. In women with a low BMI and who are diagnosed with breast cancer, nutritional programs that are specifically designed to increase nutritional intake, for example through increasing caloric intake, in order to increase BMI may be implemented. On the other hand, in women with a high BMI and who are diagnosed with breast cancer, nutritional programs that lower nutritional intake may be implemented to lower BMI. Greater emphasis should be placed on those prognostic factors that are modifiable such as BMI in order to increase survival or improve prognosis.

REFERENCES

1. Canadian Cancer Statistics 2010. Toronto: Canadian Cancer Society; 2010.
2. Hunt K, Robb G, Storm E, Ueno N. *Breast Cancer* Houston Springer-Verlag; 2001.
3. ACS. Breast Cancer Facts & Figures 2009-2010. Atlanta American Cancer Society, Inc
4. Whiteman MK, Hillis SD, Curtis KM, McDonald JA, Wingo PA, Marchbanks PA. Body mass and mortality after breast cancer diagnosis. *Cancer Epidemiol Biomarkers Prev.* Aug 2005;14(8):2009-2014.
5. Visscher TL, Seidell JC. The public health impact of obesity. *Annu Rev Public Health.* 2001;22:355-375.
6. Luo W, Morrison H, de Groh M, et al. The burden of adult obesity in Canada. *Chronic Dis Can.* 2007;27(4):135-144.
7. Trentham-Dietz A, Newcomb PA, Storer BE, et al. Body size and risk of breast cancer. *Am J Epidemiol.* Jun 1 1997;145(11):1011-1019.
8. Dawood S, Broglio K, Gonzalez-Angulo AM, et al. Prognostic value of body mass index in locally advanced breast cancer. *Clin Cancer Res.* Mar 15 2008;14(6):1718-1725.
9. Kroenke CH, Chen WY, Rosner B, Holmes MD. Weight, weight gain, and survival after breast cancer diagnosis. *J Clin Oncol.* Mar 1 2005;23(7):1370-1378.
10. Newman SC, Lees AW, Jenkins HJ. The effect of body mass index and oestrogen receptor level on survival of breast cancer patients. *Int J Epidemiol.* Jun 1997;26(3):484-490.
11. Greenberg ER, Vessey MP, McPherson K, Doll R, Yeates D. Body size and survival in premenopausal breast cancer. *Br J Cancer.* May 1985;51(5):691-697.
12. Katoh A, Watzlaf VJ, D'Amico F. An examination of obesity and breast cancer survival in post-menopausal women. *Br J Cancer.* Nov 1994;70(5):928-933.
13. Menon KV, Hodge A, Houghton J, Bates T. Body mass index, height and cumulative menstrual cycles at the time of diagnosis are not risk factors for poor outcome in breast cancer. *Breast.* Dec 1999;8(6):328-333.
14. Carmichael AR, Bendall S, Lockerbie L, Prescott RJ, Bates T. Does obesity compromise survival in women with breast cancer? *Breast.* Apr 2004;13(2):93-96.
15. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
16. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating* New York: Springer 2009.
17. NCI. What is cancer? http://www.cancer.gov/cancertopics/what-is-cancer. Accessed December 5, 2009.
18. Rose DP. Diet, hormones, and cancer. *Annu Rev Public Health.* 1993;14:1-17.
19. Cancer Reference Information: What is Breast Cancer? http://www.cancer.org/docroot/CRI/content/CRI_2_2_1X_What_is_breast_cancer_5.asp?sitearea=. Accessed November 20, 2009.

20. Chappell J, Leitner JW, Solomon S, Golovchenko I, Goalstone ML, Draznin B. Effect of insulin on cell cycle progression in MCF-7 breast cancer cells. Direct and potentiating influence. *J Biol Chem.* Oct 12 2001;276(41):38023-38028.
21. Bhalla V, Joshi K, Vohra H, Singh G, Ganguly NK. Effect of growth factors on proliferation of normal, borderline, and malignant breast epithelial cells. *Exp Mol Pathol.* Apr 2000;68(2):124-132.
22. Chang S, Alderfer JR, Asmar L, Buzdar AU. Inflammatory breast cancer survival: the role of obesity and menopausal status at diagnosis. *Breast Cancer Res Treat.* Nov 2000;64(2):157-163.
23. Singletary SE, Connolly JL. Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual. *CA Cancer J Clin.* Jan-Feb 2006;56(1):37-47; quiz 50-31.
24. Sultana T, Chowdhury AA, Rahman Q, Ahmed ANN, Rahman AJEN. A single pathological grading system for breast carcinoma should be adopted in Bangladesh. *Bangladesh J Pathology* 2009;24(2):11-14.
25. Garcia M, Jemal A, Ward E, et al. *Global Cancer Facts & Figures 2007.* Atlanta, GA: American Cancer Society; 2007.
26. Colditz GA, Sellers TA, Trapido E. Epidemiology - identifying the causes and preventability of cancer? *Nat Rev Cancer.* Jan 2006;6(1):75-83.
27. Mettlin C. Global breast cancer mortality statistics *A Cancer Journal for Clinicians* 1999;49:138-144.
28. Ryerson AB, Miller JW, Eheman CR, Leadbetter S, White MC. Recent trends in U.S. mammography use from 2000-2006: a population-based analysis. *Prev Med.* Nov 2008;47(5):477-482.
29. Breen N, K AC, Meissner HI, et al. Reported drop in mammography : is this cause for concern? *Cancer.* Jun 15 2007;109(12):2405-2409.
30. *Cancer Facts & Figures 2009.* Atlanta: American Cancer Society; 2009.
31. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med.* Oct 27 2005;353(17):1784-1792.
32. *Ontario Cancer Facts: Breast Cancer Incidence Stable in Ontario, Mortality Declining* Toronto: Cancer Care Ontario October, 2007.
33. Colditz GA, Rosner B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol.* Nov 15 2000;152(10):950-964.
34. Hartmann LC, Sellers TA, Frost MH, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med.* Jul 21 2005;353(3):229-237.
35. Steenland K, Whelan E, Deddens J, Stayner L, Ward E. Ethylene oxide and breast cancer incidence in a cohort study of 7576 women (United States). *Cancer Causes Control.* Aug 2003;14(6):531-539.
36. Huang Z, Hankinson SE, Colditz GA, et al. Dual effects of weight and weight gain on breast cancer risk. *JAMA.* Nov 5 1997;278(17):1407-1411.
37. Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D. Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ.* Dec 1 2007;335(7630):1134.

38. Sherman B, Wallace R, Bean J, Schlabaugh L. Relationship of body weight to menarcheal and menopausal age: implications for breast cancer risk. *J Clin Endocrinol Metab.* Mar 1981;52(3):488-493.
39. Rich-Edwards JW, Goldman MB, Willett WC, et al. Adolescent body mass index and infertility caused by ovulatory disorder. *Am J Obstet Gynecol.* Jul 1994;171(1):171-177.
40. Stephenson GD, Rose DP. Breast cancer and obesity: an update. *Nutr Cancer.* 2003;45(1):1-16.
41. Cleveland RJ, Eng SM, Abrahamson PE, et al. Weight gain prior to diagnosis and survival from breast cancer. *Cancer Epidemiol Biomarkers Prev.* Sep 2007;16(9):1803-1811.
42. Dignam JJ, Wieand K, Johnson KA, Fisher B, Xu L, Mamounas EP. Obesity, tamoxifen use, and outcomes in women with estrogen receptor-positive early-stage breast cancer. *J Natl Cancer Inst.* Oct 1 2003;95(19):1467-1476.
43. Toniolo PG. Endogenous estrogens and breast cancer risk: the case for prospective cohort studies. *Environ Health Perspect.* Apr 1997;105 Suppl 3:587-592.
44. McTiernan A, Rajan KB, Tworoger SS, et al. Adiposity and sex hormones in postmenopausal breast cancer survivors. *J Clin Oncol.* May 15 2003;21(10):1961-1966.
45. Enger SM, Greif JM, Polikoff J, Press M. Body weight correlates with mortality in early-stage breast cancer. *Arch Surg.* Sep 2004;139(9):954-958; discussion 958-960.
46. Caan BJ, Emond JA, Natarajan L, et al. Post-diagnosis weight gain and breast cancer recurrence in women with early stage breast cancer. *Breast Cancer Res Treat.* Sep 2006;99(1):47-57.
47. Tao MH, Shu XO, Ruan ZX, Gao YT, Zheng W. Association of overweight with breast cancer survival. *Am J Epidemiol.* Jan 15 2006;163(2):101-107.
48. Morimoto LM, White E, Chen Z, et al. Obesity, body size, and risk of postmenopausal breast cancer: the Women's Health Initiative (United States). *Cancer Causes Control.* Oct 2002;13(8):741-751.
49. Judd HL, Shamonki IM, Frumar AM, Lagasse LD. Origin of serum estradiol in postmenopausal women. *Obstet Gynecol.* Jun 1982;59(6):680-686.
50. Newcomb PA, Klein R, Klein BE, et al. Association of dietary and life-style factors with sex hormones in postmenopausal women. *Epidemiology.* May 1995;6(3):318-321.
51. Petrelli JM, Calle EE, Rodriguez C, Thun MJ. Body mass index, height, and postmenopausal breast cancer mortality in a prospective cohort of US women. *Cancer Causes Control.* May 2002;13(4):325-332.
52. Ota DM, Jones LA, Jackson GL, Jackson PM, Kemp K, Bauman D. Obesity, non-protein-bound estradiol levels, and distribution of estradiol in the sera of breast cancer patients. *Cancer.* Feb 1 1986;57(3):558-562.
53. Cauley JA, Gutai JP, Kuller LH, LeDonne D, Powell JG. The epidemiology of serum sex hormones in postmenopausal women. *Am J Epidemiol.* Jun 1989;129(6):1120-1131.

54. Del Giudice ME, Fantus IG, Ezzat S, McKeown-Eyssen G, Page D, Goodwin PJ. Insulin and related factors in premenopausal breast cancer risk. *Breast Cancer Res Treat.* Jan 1998;47(2):111-120.
55. Li BD, Khosravi MJ, Berkel HJ, et al. Free insulin-like growth factor-I and breast cancer risk. *Int J Cancer.* Mar 1 2001;91(5):736-739.
56. Rose DP, Gilhooly EM, Nixon DW. Adverse effects of obesity on breast cancer prognosis, and the biological actions of leptin (review). *Int J Oncol.* Dec 2002;21(6):1285-1292.
57. Laud K, Gourdou I, Pessemesse L, Peyrat JP, Djiane J. Identification of leptin receptors in human breast cancer: functional activity in the T47-D breast cancer cell line. *Mol Cell Endocrinol.* Feb 25 2002;188(1-2):219-226.
58. Dieudonne MN, Machinal-Quelin F, Serazin-Leroy V, Leneveu MC, Pecquery R, Giudicelli Y. Leptin mediates a proliferative response in human MCF7 breast c ancer cells. *Biochem Biophys Res Commun.* Apr 26 2002;293(1):622-628.
59. Castellucci M, De Matteis R, Meisser A, et al. Leptin modulates extracellular matrix molecules and metalloproteinases: possible implications for trophoblast invasion. *Mol Hum Reprod.* Oct 2000;6(10):951-958.
60. Tessitore L, Vizio B, Jenkins O, et al. Leptin expression in colorectal and breast cancer patients. *Int J Mol Med.* Apr 2000;5(4):421-426.
61. Moon HG, Han W, Noh DY. Underweight and breast cancer recurrence and death: a report from the Korean Breast Cancer Society. *J Clin Oncol.* Dec 10 2009;27(35):5899-5905.
62. Tammemagi CM, Nerenz D, Neslund-Dudas C, Feldkamp C, Nathanson D. Comorbidity and survival disparities among black and white patients with breast cancer. *JAMA.* Oct 12 2005;294(14):1765-1772.
63. Canadian Cancer Statistics 2009. Toronto: Canadian Cancer Society; 2009.
64. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology *Natl Cancer Inst Monogr.* 1961;6:101-121.
65. NCI. SEER Stat Fact Sheets: Breast. http://seer.cancer.gov/statfacts/html/breast.html#survival. Accessed June 2, 2010.
66. Shek LL, Godolphin W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. *Cancer Res.* Oct 1 1988;48(19):5565-5569.
67. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models.* New York, NY: Springer 2005.
68. Caan BJ, Kwan ML, Hartzell G, et al. Pre-diagnosis body mass index, post-diagnosis weight change, and prognosis among women with early stage breast cancer. *Cancer Causes Control.* Dec 2008;19(10):1319-1328.
69. Parl FF, Schmidt BP, Dupont WD, Wagner RK. Prognostic significance of estrogen receptor status in breast cancer in relation to tumor stage, axillary node metastasis, and histopathologic grading. *Cancer.* Nov 15 1984;54(10):2237-2242.
70. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. *Br J Cancer.* Mar 1982;45(3):361-366.

71. Goodwin PJ, Ennis M, Pritchard KI, et al. Fasting insulin and outcome in early-stage breast cancer: results of a prospective cohort study. *J Clin Oncol.* Jan 1 2002;20(1):42-51.
72. Kushi LH, Byers T, Doyle C, et al. American Cancer Society Guidelines on Nutrition and Physical Activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J Clin.* Sep-Oct 2006;56(5):254-281; quiz 313-254.
73. Cummings SR, Tice JA, Bauer S, et al. Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. *J Natl Cancer Inst.* Mar 18 2009;101(6):384-398.
74. Berghofer A, Pischon T, Reinhold T, Apovian CM, Sharma AM, Willich SN. Obesity prevalence from a European perspective: a systematic review. *BMC Public Health.* 2008;8:200.
75. James PT. Obesity: the worldwide epidemic. *Clin Dermatol.* Jul-Aug 2004;22(4):276-280.
76. Katzmarzyk PT, Mason C. Prevalence of class I, II and III obesity in Canada. *CMAJ.* Jan 17 2006;174(2):156-157.
77. Jain MG, Miller AB, Rohan TE, et al. Body mass index and mortality in women: follow-up of the Canadian National Breast Screening Study cohort. *Int J Obes (Lond).* Jul 2005;29(7):792-797.
78. WHO. Global database on body mass index: BMI classification http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Accessed December 11, 2009.
79. Liu LN, Miaskowski C, Wang JS, Chen SC, Chen ML. Accuracy of body mass index to determine obesity in women with breast cancer: An observational study of Taiwanese sample. *Int J Nurs Stud.* Feb 1.
80. Loi S, Milne RL, Friedlander ML, et al. Obesity and outcomes in premenopausal and postmenopausal breast cancer. *Cancer Epidemiol Biomarkers Prev.* Jul 2005;14(7):1686-1691.
81. Zhang S, Folsom AR, Sellers TA, Kushi LH, Potter JD. Better breast cancer survival for postmenopausal women who are less overweight and eat less fat. The Iowa Women's Health Study. *Cancer.* Jul 15 1995;76(2):275-283.
82. Daling JR, Malone KE, Doody DR, Johnson LG, Gralow JR, Porter PL. Relation of body mass index to tumor markers and survival among young women with i nvasive ductal breast carcinoma. *Cancer.* Aug 15 2001;92(4):720-729.
83. Abrahamson PE, Gammon MD, Lund MJ, et al. General and abdominal obesity and survival among young women with breast cancer. *Cancer Epidemiol Biomarkers Prev.* Oct 2006;15(10):1871-1877.
84. Reeves KW, Faulkner K, Modugno F, et al. Body mass index and mortality among older breast cancer survivors in the Study of Osteoporotic Fractures. *Cancer Epidemiol Biomarkers Prev.* Jul 2007;16(7):1468-1473.
85. Carmichael AR. Obesity and prognosis of breast cancer. *Obes Rev.* Nov 2006;7(4):333-340.
86. Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. *JAMA.* Apr 20 2005;293(15):1861-1867.

87. Suissa S, Pollak M, Spitzer WO, Margolese R. Body size and breast cancer prognosis: a statistical explanation of the discrepancies. *Cancer Res.* Jun 1 1989;49(11):3113-3116.
88. Harrell FE. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. New York: Springer; 2001.
89. Dupont WD. *Statistical modeling for biomedical researchers: a simple introduction to the analysis of complex data*. 2 ed. Cambridge; New York: Cambridge University Press; 2009.
90. White E, Armstrong BK, Saracci R. *Principles of exposure measurement in epidemiology: collecting, evaluating, and improving measures of disease risk factors*. 2nd ed. Oxford; New York: Oxford University Press; 2008.
91. Gail MH, Pfeiffer RM. On criteria for evaluating models of absolute risk. *Biostatistics.* Apr 2005;6(2):227-239.
92. Kaplan D, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53:457-481.
93. Cox D. Regression models and life tables (with discussion). *J R Stat Soc.* 1972;34:187-220.
94. Cleves MA. *An introduction to survival analysis using Stata.* 2nd ed. College Station, Tex.: Stata Press; 2008.
95. Jacoby WG. Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies.* 2000;19:577–613.
96. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15(4):361-387.
97. Hosmer DW, Jr, Lemeshow S. *Applied Logistic Regression.* 2nd Edition ed. New York: John Wiley & Sons, Inc.; 1999.
98. Chen CH, Lo YF, Tsai HP, et al. Low body mass index is an independent risk factor of locoregional recurrence in women with breast cancer undergoing breast conserving therapy *Chang Gung Med J.* 2009;32(5):553-561.
99. Marret H, Perrotin F, Bougnoux P, et al. Low body mass index is an independent predictive factor of local recurrence after conservative treatment for breast cancer *Breast Cancer Res Treat.* 2001;66:17-23.
100. Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Stat Methods Med Res.* Jun 2007;16(3):277-298.
101. Beenken SW, Grizzle WE, Crowe DR, et al. Molecular biomarkers for breast cancer prognosis: coexpression of c-erbB-2 and p53. *Ann Surg.* May 2001;233(5):630-638.
102. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol.* Feb 1998;11(2):155-168.
103. Lai H, Ma F, Trapido E, Meng L, Lai S. Spectrum of p53 tumor suppressor gene mutations and breast cancer survival. *Breast Cancer Res Treat.* Jan 2004;83(1):57-66.
104. Gordis L. *Epidemiology* 4ed. Philadelphia: Elsevier Inc; 2009.

105. Kumar NB, Lyman GH, Allen K, Cox CE, Schapira DV. Timing of weight gain and breast cancer risk. *Cancer.* Jul 15 1995;76(2):243-249.

APPENDICES

Appendix 1- Graphical and Statistical Assessment of Proportional Hazards Assumptions for Cox Proportional Hazards Model

Table of Contents
Breast Cancer-Specific Death

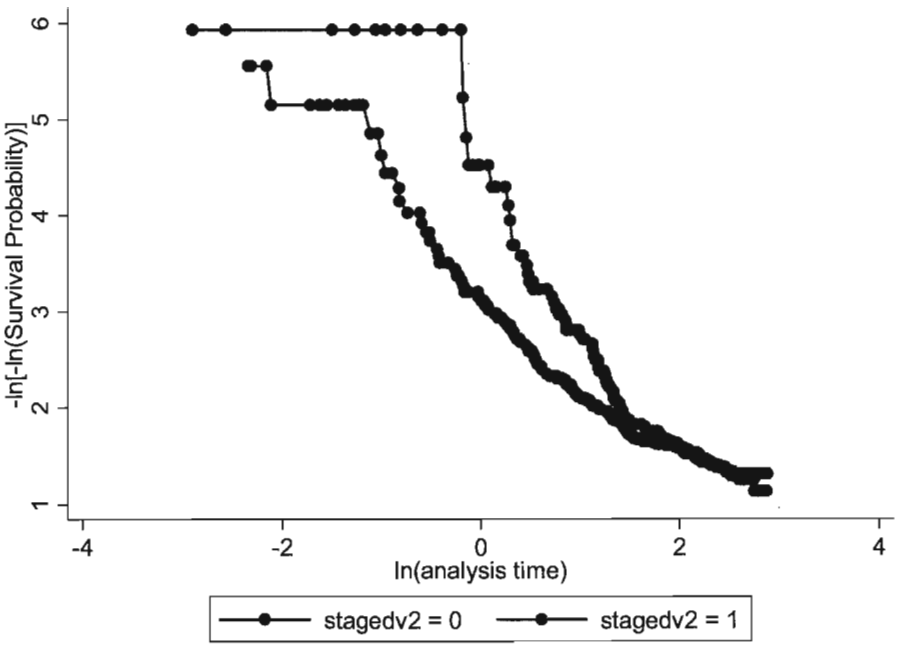


***Figure 1.1*- Graphical Test of Proportional Hazards Assumption between Stage II and Breast Cancer-Specific Death**

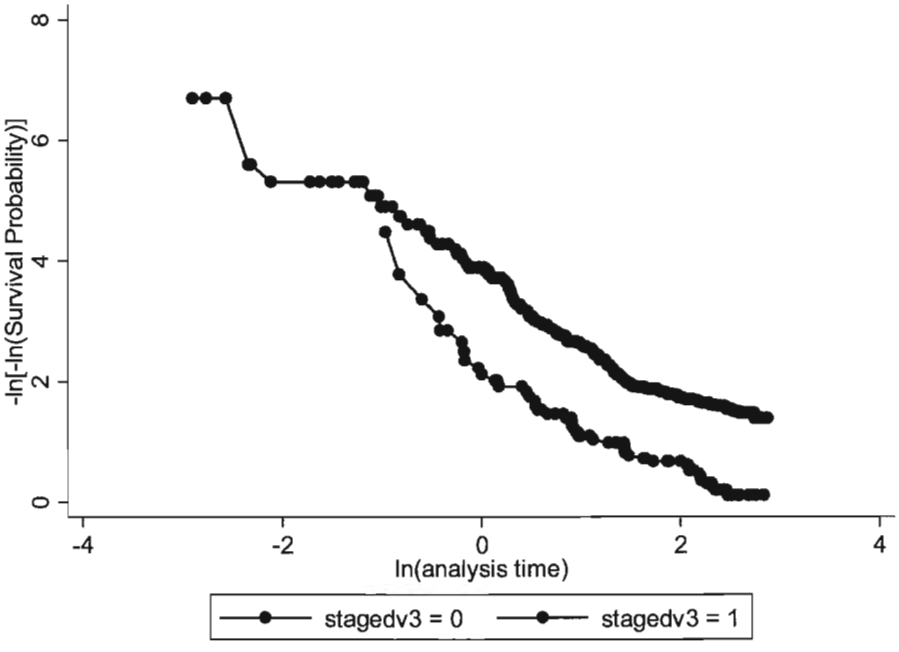


***Figure 1.2*- Graphical Test of Proportional Hazards Assumption between Stage III and Breast Cancer-Specific Death**

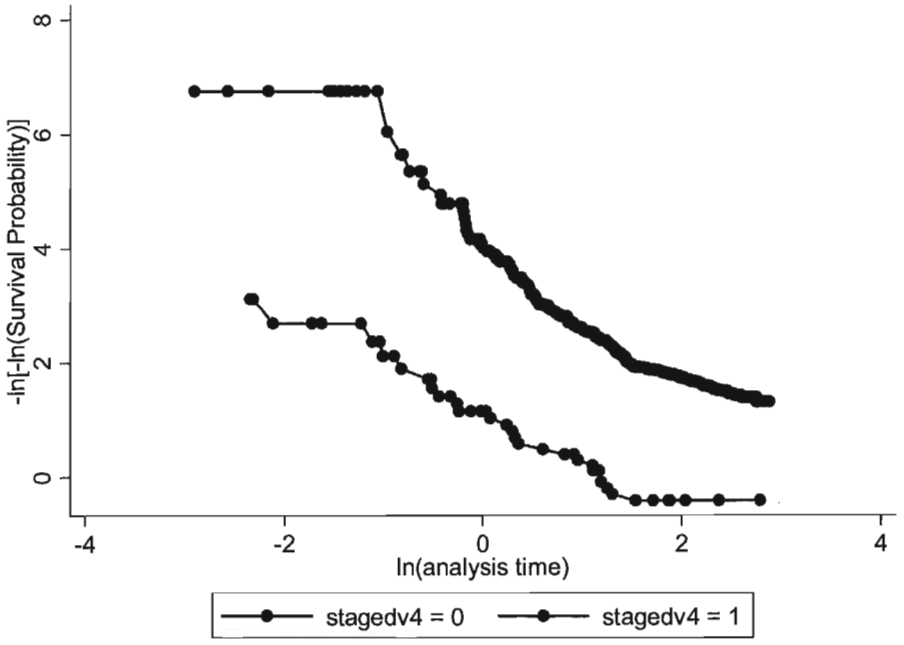


***Figure 1.3*- Graphical Test of Proportional Hazards Assumption between Stage IV and Breast Cancer-Specific Death**

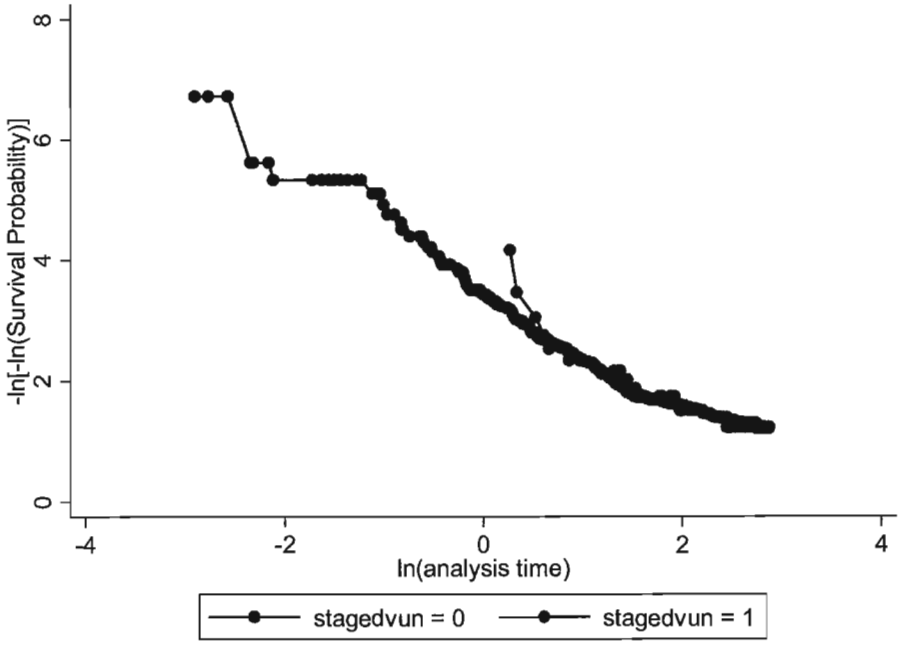


***Figure 1.4*- Graphical Test of Proportional Hazards Assumption between Stage Unknown and Breast Cancer-Specific Death**

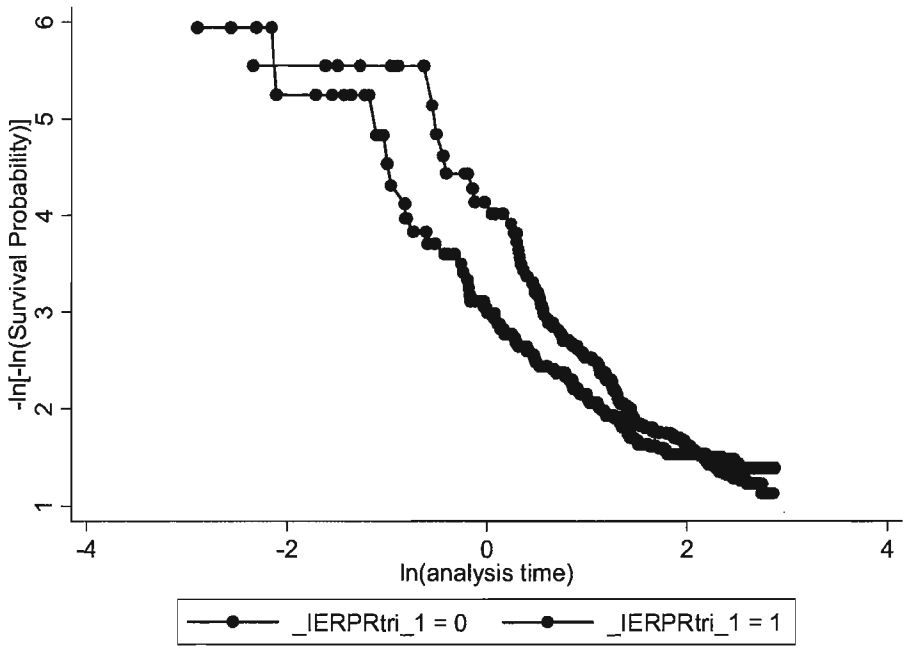


***Figure 1.5*- Graphical Test of Proportional Hazards Assumption between ER/PR + and Breast Cancer-Specific Death**

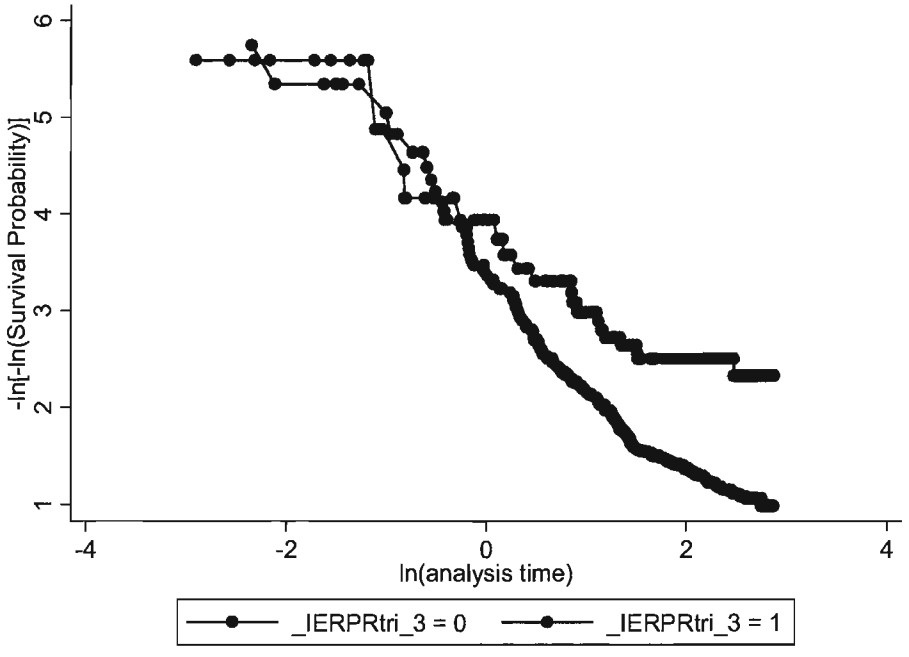


***Figure 1.6*- Graphical Test of Proportional Hazards Assumption between ER/PR Unknown and Breast Cancer-Specific Death**

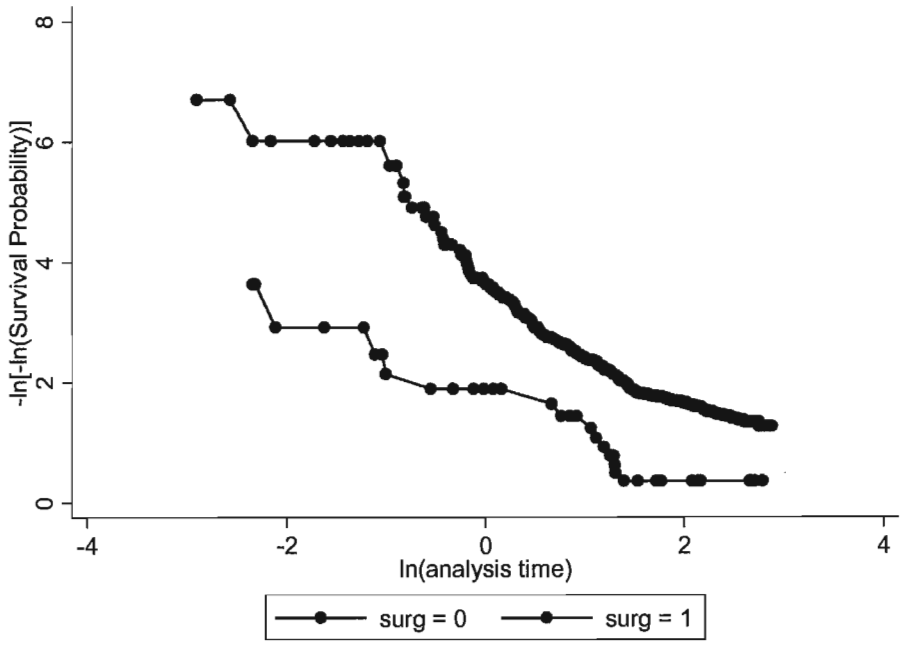


***Figure 1.7*- Graphical Test of Proportional Hazards Assumption between Surgery and Breast Cancer-Specific Death**

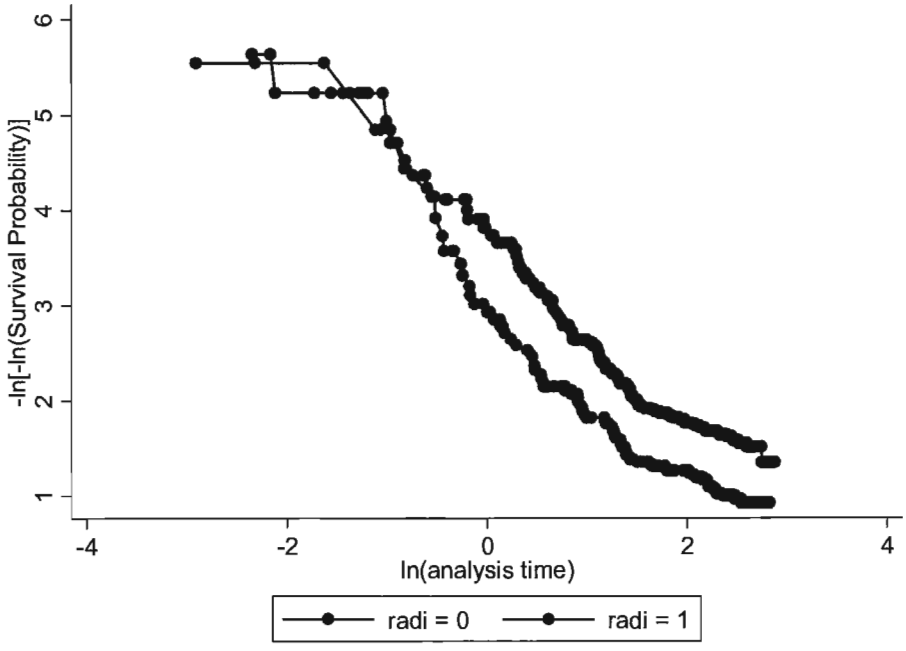


***Figure 1.8*- Graphical Test of Proportional Hazards Assumption between Radiation Therapy and Breast Cancer-Specific Death**

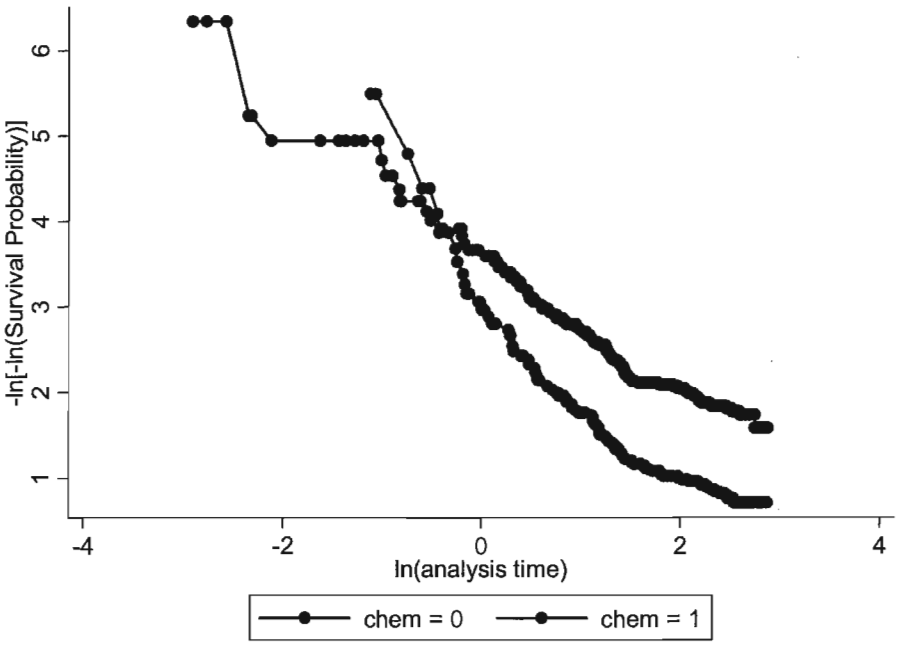


***Figure 1.9*- Graphical Test of Proportional Hazards Assumption between Chemotherapy and Breast Cancer-Specific Death**

***Table 1.1*- Test of Proportional Hazards Assumption for Model Predicting Breast Cancer-Specific Death**

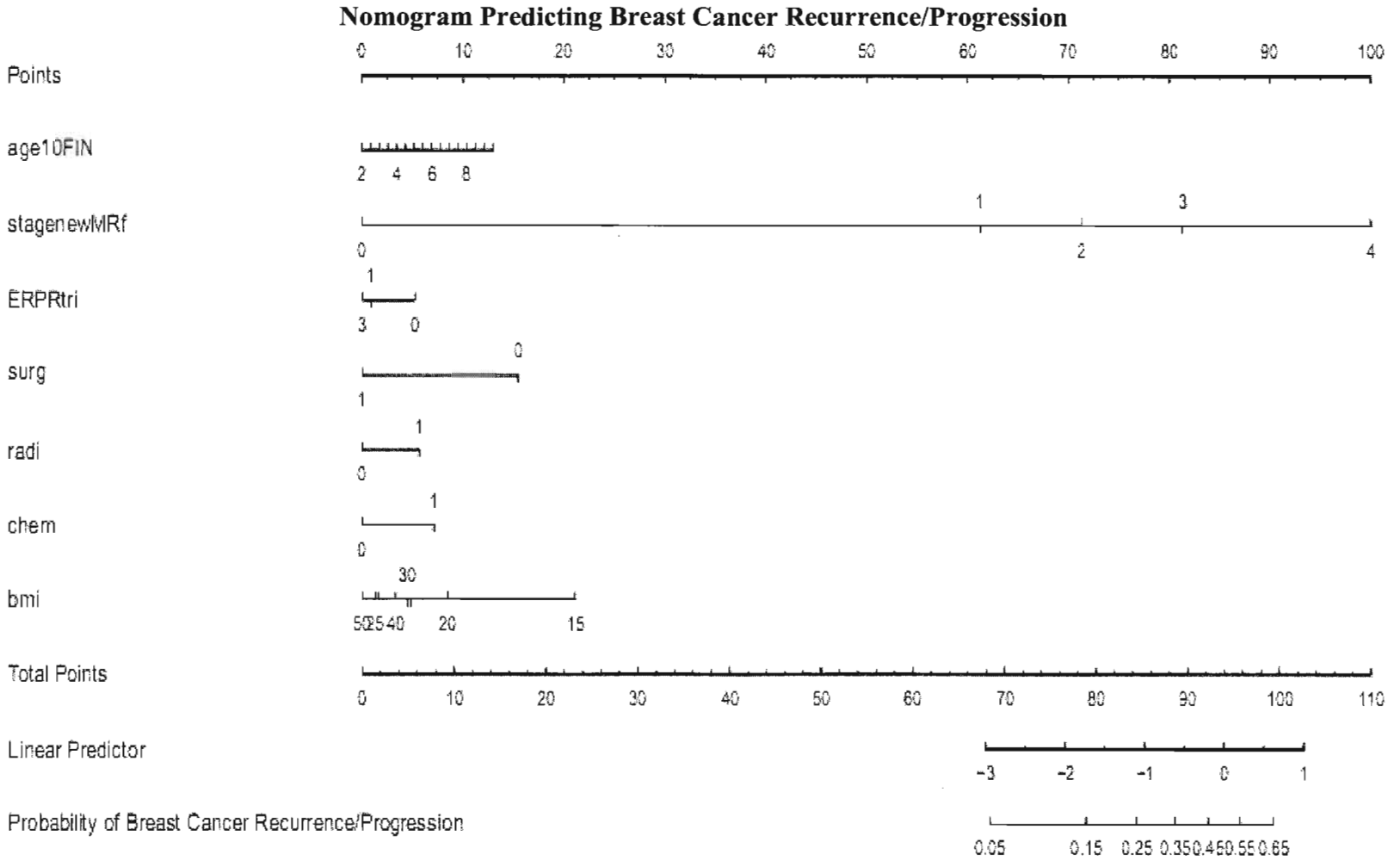
Time: Rank(t)

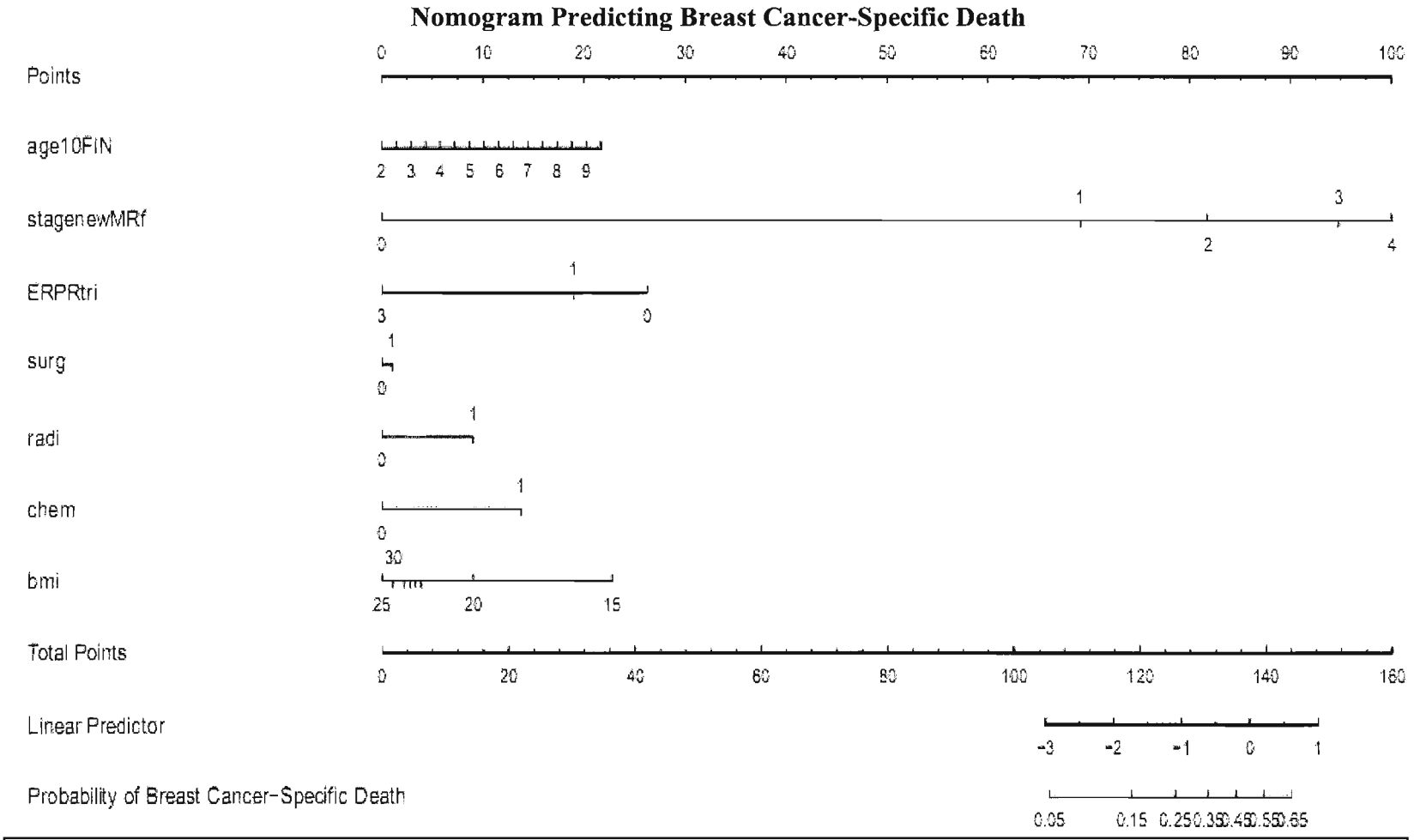
	rho	chi2	df	Prob>chi2
age10FIN	0.04777	0.39	1	0.5329
stagedv2	0.05810	0.51	1	0.4769
stagedv3	0.02024	0.07	1	0.7964
stagedv4	-0.07410	0.99	1	0.3205
stagedvun	0.02293	0.07	1	0.7869
_IERPRtri_1	0.32676	17.90	1	0.0000
_IERPRtri_3	0.02301	0.08	1	0.7789
surg	-0.05643	0.63	1	0.4288
radi	0.06198	0.60	1	0.4390
chem	0.17511	5.64	1	0.0176
BMI4k3sp1	-0.14671	4.03	1	0.0448
BMI4k3sp2	0.15921	4.35	1	0.0370
BMI4k3sp3	-0.15529	4.00	1	0.0454
global test		31.22	13	0.0031

Appendix 2- Missing Value Patterns

Variable	type	obs	mv	variable label
BRCArecprog2	float	799	98	Recurrence/progression
expireBRCA	float	897	0	Breast cancer death
age10FIN	double	897	0	
stagedv2	float	897	0	
stagedv3	float	897	0	
stagedv4	float	897	0	
stagedvun	float	897	0	
_IERPRtri_1	byte	897	0	ER/PR positive
_IERPRtri_3	byte	897	0	ER/PR negative
surg	byte	858	39	
radi	byte	819	78	
chem	byte	811	86	
BMI4k3sp1	float	816	81	
BMI4k3sp2	float	816	81	
BMI4k3sp3	float	816	81	

Appendix 3- Nomogram





Directions: To obtain the nomogram predicted probability, locate the patient values at the age, stage, ER/PR status, treatment (surgery, radiation therapy and chemotherapy) and BMI axes. Then, draw a vertical line to the "points" axis to determine how many points are attributed for each variable value. Sum the points for all the variables described above. Locate the sum on the "total points" line and draw a vertical line to the appropriate probability scale, to obtain the predicted probability for the breast cancer outcomes.